AI Hub Data Commons Collaborative Consulting Services

RFP No. 2026-JAII-01
IBM RFP RESPONSE PREPARED FOR: Massachusetts Technology Collaborative

— October 7, 2025

Submitted to:
Massachusetts Technology Collaborative
75 North Drive
Westborough, MA 01581

Submitted by:
IBM CORPORATION
Shawn Noyes
75 Binney St, Cambridge, MA, 02142

617.901.5848

“Whenever an individual or a business decides that success has been attained, progress stops.”

Thomas J. Watson
Former Chairman and CEO of IBM

Table of Contents

Attachment A ***4***

Attachment B ***5***

Attachment C ***6***

1. Executive Summary ***10***

2. Our Solution ***12***

3. IBM Expert Labs Implementation Services ***28***

4. Integration with Chosen SI's PMO ***35***

5. IBM Public Sector References ***37***

6. Conclusion ***41***

Appendix – Pricing ***43***

Appendix – Description of Services ***47***

IBM W9 ..... ***61***

Attachment A

Application Cover Sheet

<table>
<tr><td colspan="4">Name of Respondent: Shawn Noyes, IBM Corporation</td></tr>
<tr><td>Mailing Address:</td><td>75 Binney Street, Cambridge, Massachusetts, 02142</td><td>Massachusetts</td><td>02142</td></tr>
<tr><td>Telephone: 617.901.5848</td><td>Fax:</td><td colspan="2">Web Address: https://www.IBM.com</td></tr>
<tr><td colspan="2">Primary Contact for Clarification: Shawn Noyes</td><td colspan="2">Primary Contact E-mail Address: shawn.Noyes@IBM.com</td></tr>
<tr><td colspan="2">Authorized Signatory: Shawn Noyes</td><td colspan="2">Authorized Signatory E-mail Address: Shawn.Noyes@IBM.com</td></tr>
<tr><td colspan="2">Legal Status / Jurisdiction: United States</td><td colspan="2">Respondents W-9: W9 document attached to submission
Respondents UEI No.: 23-0871985</td></tr>
</table>

Attachment B

Massachusetts Technology Collaborative RFP No. 2026-JAII-01 - Attachment B Massachusetts Technology Collaborative

Authorized Respondent's Signature and Acceptance Form

The undersigned is a duly authorized representative of the Respondent listed below. The Respondent has read and understands the RFP requirements. The Respondent acknowledges that all of the terms and conditions of the RFP are mandatory, and that Respondent's response is compliant with such requirements. The Respondent understands that, if selected by the Mass Tech Collaborative, the Respondent and the Mass Tech Collaborative will execute an Agreement specifying the mutual requirements of participation. Respondent agrees that the entire bid response will remain valid for sixty (60) days from receipt by the Mass Tech Collaborative. I certify that Respondent is in compliance with all corporate filing requirements and State tax laws. I certify that the statements made in this response to the RFP, including all attachments and exhibits, are true and correct to the best of my knowledge. I acknowledge and understand the procedures for handling confidential materials submitted to the Mass Tech Collaborative as set forth in Section 3.1(d); I agree to be bound by those procedures and agree that the Mass Tech Collaborative shall not be liable under any circumstances for the disclosure of any materials submitted to the Mass Tech Collaborative pursuant to this RFP or upon the Respondent's selection.

Respondent: Shawn M. Noyes

By: *Shawn M. Noyes*

Shawn M. Noyes

07 October 2025

Shawn.Noyes@IBM.com, Cell: 617.901.5848, IBM Technology Sales Leader, Commonwealth of Massachusetts

Attachment C

Massachusetts Technology Collaborative RFP No. 2026-JAII-01

Attachment C Budget – Customer and IBM understand that this may be included into a larger proposal via a systems integrator. Customer may procure these services directly, too.

Software - IBM Technology							
Year 1				Year 2		Year 3	
Description	SKU	Quantity	Investement	Quantity	Investement	Quantity	Investement
Watsonx.data Intelligence	D10LXZX	2,500	$ 165,600.00	2,500	$ 182,160.00	2,500	$ 200,376.00
Watsonx.data	D0HLLZX	9	$ 8,280.00	9	$ 9,108.00	9	$ 10,018.80
Cloud Pak for Data - PostgressSQL	D01CFZX	200	$ 2,400.00	200	$ 2,640.00	200	$ 2,904.00
watsonx.gov	D0FHQZX	6,750	$ 74,520.00	6,750	$ 81,972.00	6,750	$ 90,169.20
watsonx.ai (synthetic data generation)	D0FHQZX	167	$ 1,840.00	167	$ 2,024.00	167	$ 2,226.40
watsonx.orchestrate	D0B3YZX	167	$ 5,520.00	167	$ 6,072.00	167	$ 6,679.20
Apptio Standard	D0ICKZX	15	$ 118,800.00	15	$ 130,680.00	15	$ 143,748.00
Apptio All Education Pass Quantity	D0GWJZX	2	$ 7,704.00	2	$ 8,474.40	2	$ 9,321.84
IBM Cloud Pak for Integration - API Calls Resource Value Unit	D052XZX	30	$ 16,632.00	150	$ 83,160.00	150	$ 91,476.00
Year 1 Software Total			**$ 401,296.00**		**$ 506,290.40**		**$ 556,919.44**
Expert Labs - IBM Services							
Year 1				**Year 2**		**Year 3**	
Expert Labs- Data Catalog Build	D12DIZX	1	$40,000.00				
Expert Labs- Data Quality Build	D10DTZX	1	$50,000.00				
Expert Labs- Lineage Build	D10DTZX	1	$50,000.00				
IBM Expert Labs AI Governance One Week Remote Solution Build - SaaS Engagement Remotely Delivered	D070BZX	4	$ 40,000.00				
IBM Expert Labs Build AI Governance - SaaS Engagement Remotely Delivered SW Services	D0GC7ZX	5	$ 86,500.00				
IBM Expert Labs Advise AI Governance - SaaS Engagement Remotely Delivered SW Services	D0GC9ZX	2	$ 38,200.00				
IBM Expert Labs Advise AI Assistants SaaS Engagement Remotely Delivered Service	D06ZVZX	2	$ 38,200.00				
IBM Expert Labs Build AI Assistants SaaS Engagement Remotely	D06ZWZX	3	$ 51,900.00				
Expert Labs Implementation Quantity	D0IG9ZX	2	$ 158,000.00				
Expert Labs Managed Services Quantity	D0HGAZX	1	$ 52,800.00				
IBM Expert Labs Install Integration - Non-SaaS Engagement Remotely Delivered SW Services	D02DAZX	3	$ 51,900.00				
IBM Exp Labs - IBM Cloud Pak for Integration One Week Soln Arch-Non-SaaS Engagement Remotely Delivered SW Services	D02DBZX	2	$ 45,000.00				
IBM Expert Labs Build Integration - Non-SaaS Engagement Remotely Delivered SW Services	D02DCZX	4	$ 69,200.00				
Year 1 Investment			**$1,172,996.00**	**Year 2 Investment**	**$ 506,290.40**	**Year 3 Investment**	**$ 556,919.44**

October 7, 2025

Massachusetts Technology Collaborative
75 North Drive
Westborough, MA 01581

Subject: Response to RFP No. 2026-JAII-01 – AI Hub Data Commons Collaborative Consulting Services

Dear Selection Committee,

On behalf of IBM Technology, we are thrilled to submit our proposal to support the Massachusetts Technology Collaborative AI Hub (MassTech) in advancing the Data Commons Collaborative initiative. IBM is committed to helping the Commonwealth of Massachusetts realize its vision for a secure, scalable, and innovative data platform that empowers research, collaboration, and responsible AI development.

We recognize the complexity and ambition of the solution outlined in RFP No. 2026-JAII-01. The requirements, as detailed in the RFP and clarified through the Q&A, demand a robust technical architecture, rigorous governance, seamless interoperability, and a scalable, secure platform to support advanced AI and data innovation across the Commonwealth. Delivering this vision requires technical excellence, strategic planning, stakeholder engagement, and a phased implementation approach.

To meet these needs, IBM has developed a comprehensive scope, objectives, and approach that addresses key functional and technical requirements for the Data Commons Collaborative. Our phased implementation strategy ensures incremental delivery of capabilities, continuous validation against MassTech's objectives, and flexibility to adapt to evolving needs—building on a solid foundation to minimize risks and maximize value. IBM Expert Labs will collaborate with the selected System Integrator to refine requirements, validate technical architecture, and align stakeholder expectations, establishing a prioritized, actionable roadmap informed by industry best practices and real-world constraints.

In collaboration with selected System Integrators (SIs) and data platform provider, IBM offers a future-ready architecture tailored to the RFP's requirements. IBM Expert Labs provides specialized implementation expertise, while IBM Technology delivers cutting-edge tools and platforms, integrated with 3rd party cloud-native data solutions (e.g., Snowflake), to ensure scalability and seamless integration.

Our team, comprising highly skilled professionals from IBM Technology and IBM Expert Labs, has deep expertise in hybrid cloud, AI, data engineering, and security. With a strong presence in Massachusetts and a proven track record of successful public sector engagements, IBM is well-positioned to deliver rapid, high-quality results. We are committed to collaborating closely with MassTech to develop an innovative and sustainable solution.

Given the complexity of the Data Commons Collaborative, IBM recommends initiating the engagement with a comprehensive design phase to refine functional requirements, validate the technical architecture, and engage stakeholders to align with MassTech's strategic goals. This phase will enable the development of a prioritized, actionable implementation plan informed by real-world constraints and best practices, laying the foundation for a successful transformation.

2025 IBM CORPORATION | CLIENT AND IBM CONFIDENTIAL

IBM, in partnership with selected SIs, are committed to delivering a solution that meets current needs and adapts to future opportunities. We are eager to collaborate with MassTech to advance the Commonwealth's leadership in AI and data innovation.

Thank you for considering our proposal. Please contact us for any additional information or clarification.

Shawn Noyes, IBM Technology Sales Lead, IBM, Commonwealth of Massachusetts and Higher Education

Shawn M. Noyes

Executive Summary

1. Executive Summary

Our Approach

Overview

This technology envisioned by this RFP will provide a key input to the Mass AI Hub, and will foster participation and innovation by both public and private sector stakeholders. IBM looks forward to partnering with whichever systems integrator (SI) is chosen to lead and manage the entire project. Our point of view is to present IBM's industry leading data and platform management technologies, along with our world class implementation services, to seamlessly integrate with a comprehensive solution. IBM has a proven track record of delivering innovative technology implementations for state agencies across the United States, and our contributions to the innovative objectives of MassTech will solidify the success of the Data Commons Collaborative.
The IBM Solution demonstrates key areas where IBM Technology can successfully complement the approach of a (SI) and data platform provider. Specifically, we outline a technology approach to metadata management, synthetic data generation, AI fairness capabilities, API management, and a Commons credit system.

Core to IBM's strategy is the concept of open architecture; the ability to interoperate in heterogenous environments. In addition to technical interoperability, our partnership alliances with System Integrators, ecosystem partners, and software publishers ensure our clients are well-positioned for success. The IBM solution is not a full-service project management and technical consulting services proposal, however, we are confident MassTech will discover a compelling point-of-view of key technology areas where IBM can contribute to the success of the DCC.

Therefore, as outlined by MassTech in Section 3.1 of the RFP, IBM is proposing an alternative approach to the solicitation.

Note: For the purposes of this response, we used Snowflake as an example of the data platform that will integrate with the proposed IBM technology solution.

IBM's Global and Local Presence

With a global workforce of highly skilled professionals, IBM delivers comprehensive technology solutions, including AI, data infrastructure, and cloud services, across diverse industries. In Massachusetts, IBM maintains a strong local presence with key locations such as our Lowell office, dedicated to software development and engineering excellence; our Cambridge facilities at 75 Binney Street and 314 Main Street, home to cutting-edge research and innovation; and the IBM Expert Labs teams providing specialized on-site and remote expertise. Additionally, through the IBM-MIT Innovation Center in Kendall Square, we foster groundbreaking AI advancements in partnership with MIT. This robust footprint—spanning over 2,000 professionals in the state—enables us to respond rapidly to client needs, offering deep expertise in AI, data engineering, and modern data estate design.

Commitment to Massachusetts

IBM is deeply committed to the Commonwealth of Massachusetts, with team members actively engaged in community initiatives, supporting organizations such as the Boys and Girls Club of Boston, Massachusetts Mentoring Partnership, and local educational institutions. IBM has also invested significantly in Massachusetts non-profits, reinforcing our dedication to building stronger communities.

Proven Experience in Massachusetts and Beyond

IBM has served numerous Massachusetts state agencies, academic institutions, and legislative bodies, addressed complex challenges and driven innovation. Our extensive public sector experience spans multiple U.S. states, federal agencies, and leading global enterprises, with expertise in architecture design, metadata management, security, governance, AI innovation, and fairness—key workstreams aligned with this solicitation.

Strategic Collaboration with Snowflake

Through our strategic alliance with Snowflake, IBM delivers cutting-edge solutions tailored to client needs. IBM Technology and IBM Expert Labs provide robust tools and specialized expertise, enabling us to address specific requirements with precision. Our collaboration with Snowflake enhances our capabilities in data infrastructure, engineering, and modern data estate implementation, ensuring we remain at the forefront of technological advancements to deliver innovative and responsible AI solutions.

Why IBM?

Deep Expertise in AI Governance and Data Intelligence

IBM brings unparalleled expertise in building trusted AI and data platforms that balance innovation with responsibility. As the creator of watsonx—recognized by Gartner as the industry leader in AI governance and evaluation technology—we deliver proven capabilities in metadata management, bias detection across 120+ demographic variables, and comprehensive data lineage tracking that are essential to the Data Commons' success. Our watsonx platform suite powers AI initiatives across many public sector organizations as well as Fortune 500 companies, demonstrating our ability to deliver enterprise-grade reliability, security, and performance at scale.

Proven Public Sector Track Record

IBM's extensive experience serving Massachusetts state agencies, academic institutions, and legislative bodies demonstrates our ability to navigate complex regulatory environments while delivering transformative outcomes. Our public sector expertise spans architecture design, metadata management, security, governance, AI fairness, and compliance—precisely the workstreams critical to the Data Commons Collaborative. We've successfully implemented similar data platforms for state agencies including California, Virginia, Louisiana, and federal entities, consistently delivering solutions

that meet stringent compliance requirements while enabling innovation. This proven track record, combined with our recognition as the #2 global AI Strategy Leader in IDC's MarketScape rankings, ensures MassTech receives best-in-class capabilities aligned with emerging regulations including the EU AI Act and NIST AI Risk Management Framework.

Technology Leadership and Responsible Innovation

IBM's technology solutions lead on security, reliability, trustworthiness, and cost-performance—qualities essential for a platform handling sensitive research data and supporting Massachusetts' AI research community. Our watsonx.governance platform embeds ethical AI principles into every workflow, providing automated bias detection, fairness monitoring, and compliance validation that protect against discriminatory outcomes while promoting inclusive model development. IBM API Connect's enterprise-grade security, with OAuth 2.0 enforcement, threat protection, and comprehensive audit logging, ensures that every data transaction meets the highest standards for privacy and transparency. Our strategic collaboration with Snowflake enhances these capabilities with cloud-native scalability, enabling the Data Commons to grow seamlessly as Massachusetts' AI ecosystem expands.

IBM Expert Labs Implementation Excellence

IBM Expert Labs brings specialized implementation expertise through proven service offerings including Build Data Catalog (SaaS), Build Data Quality (SaaS), Manta Data Lineage QuickStart, and Architect for API Connect—methodologies refined through hundreds of successful engagements worldwide. Our practitioners hold advanced certifications and maintain direct access to IBM's product roadmaps and engineering teams, ensuring MassTech benefits from cutting-edge capabilities and rapid issue resolution. This combination of technical depth, implementation methodology, and sustained partnership support guarantees not just initial platform delivery but long-term sustainability and continuous improvement aligned with the Commonwealth's evolving needs.

By leveraging IBM Technology, IBM Expert Labs, and strategic partnerships, we deliver the comprehensive governance intelligence, ethical AI oversight, and secure API orchestration that transform the Data Commons from infrastructure into a catalyst for responsible AI innovation—positioning Massachusetts as the national leader in trustworthy, accessible, and impactful AI research.

2. Our Solution

IBM's contribution to the Massachusetts AI Hub Data Commons Collaborative centers on four specialized pillars that complement the System Integrator's platform development and Snowflake's data infrastructure capabilities. First, we deliver **enterprise-grade governance and metadata management** through IBM watsonx.data Intelligence and IBM Manta Data Lineage, providing automated metadata extraction, comprehensive data provenance tracking, and alignment with open standards (Dublin Core, Schema.org, DCAT) to ensure datasets are findable, accessible, and trustworthy. Second, we provide **AI fairness and ethical oversight** via watsonx.governance, offering bias detection across 120+ demographic variables, automated compliance validation against regulations including HIPAA and the EU AI Act, and continuous monitoring to prevent fairness drift in deployed models. Third, we implement the **Commons Credits System** using IBM PostgreSQL for immutable ledger capabilities, IBM watsonx Orchestrate for transaction workflows, and IBM Apptio for

sophisticated analytics—all secured and orchestrated through IBM API Connect to ensure transparent, auditable resource allocation. Fourth, we supply **secure API orchestration** through IBM API Connect, which serves as the unified gateway enforcing OAuth 2.0 authentication, rate limiting, and comprehensive audit logging across all interactions between the SI's web portal, IBM's governance tools, and Snowflake's data platform.

IBM does not provide the user-facing web portal, frontend interface design, or primary system integration—these critical capabilities remain with the selected System Integrator, who will leverage our backend APIs to create intuitive user experiences. Similarly, while we catalog and provide synthetic data generation tools through watsonx.ai, we rely on Snowflake for core data storage, processing, and compute infrastructure, ensuring our governance layer integrates seamlessly with their scalable cloud-native architecture. This clear delineation of responsibilities—IBM for governance intelligence, the SI for user experience, and Snowflake for data operations—creates a cohesive ecosystem where each partner contributes specialized expertise toward MassTech's vision of a trusted, scalable platform that accelerates responsible AI innovation across the Commonwealth.

2.1 Overall System Architecture

The DCC architecture embodies three core principles:

1. Separation of Concerns: IBM focuses on governance, metadata management, and AI trust; the SI manages system integration and user experience; Snowflake provides the data processing foundation
2. API-First Integration: All components communicate through secure, standardized REST APIs orchestrated by IBM API Connect
3. Progressive Delivery: Ensures validated functionality at each milestone

The architecture serves four primary user communities:

1. Researchers and Academics: Require curated datasets, synthetic data generation, and compute resources for AI development
2. Government Agencies: Need secure data sharing with robust governance and compliance frameworks
3. Private Sector Innovators: Seek access to datasets and AI tools while maintaining commercial confidentiality
4. Platform Administrators: Manage credits, metadata, compliance, and system health

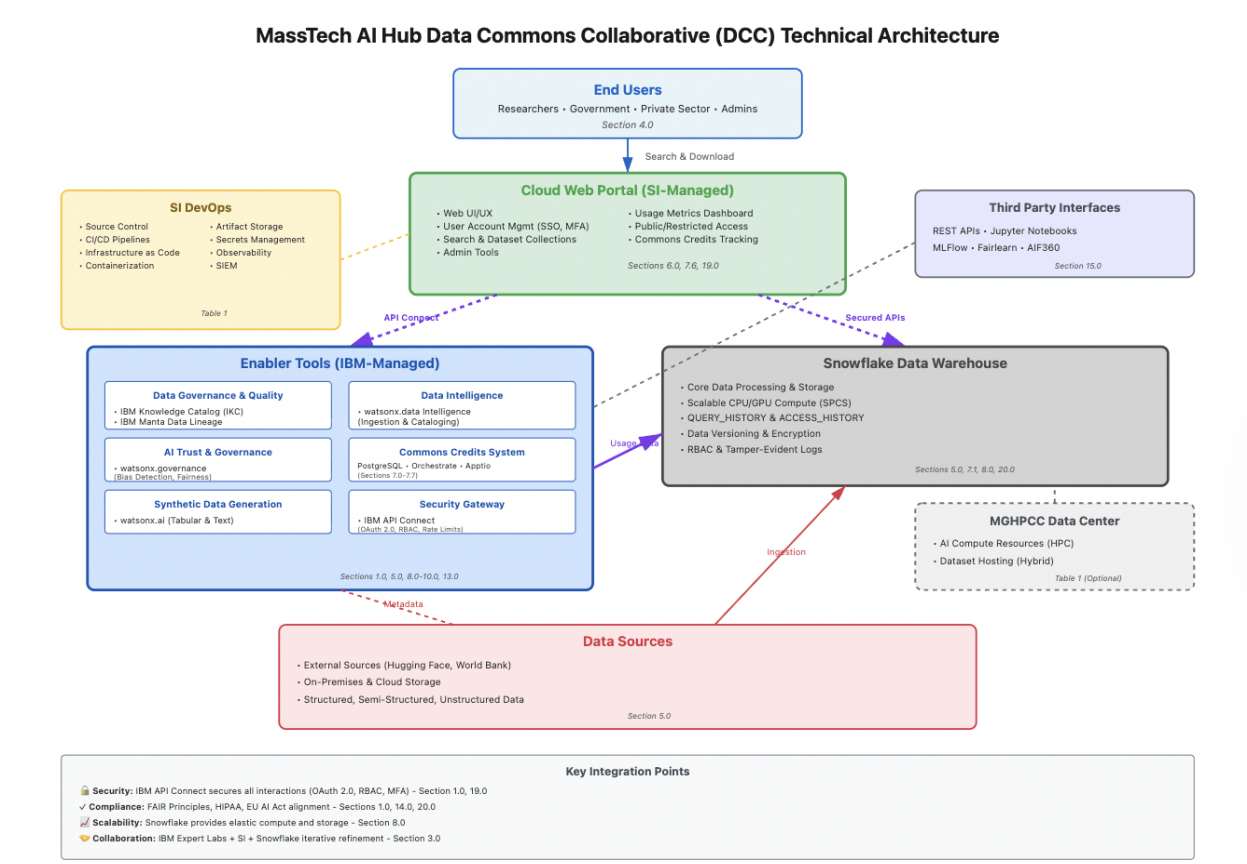

MassTech AI Hub Data Commons Collaborative (DCC) Technical Architecture
End Users
Researchers • Government • Private Sector • Admins
Section 4.0
Search & Download
Cloud Web Portal (SI-Managed)
• Web UI/UX
• User Account Mgmt (SSO, MFA)
• Search & Dataset Collections
• Admin Tools
• Usage Metrics Dashboard
• Public/Restricted Access
• Commons Credits Tracking
Sections 6.0, 7.6, 19.0
SI DevOps
• Source Control
• CI/CD Pipelines
• Infrastructure as Code
• Containerization
• Artifact Storage
• Secrets Management
• Observability
• SIEM
Table 1
Third Party Interfaces
REST APIs • Jupyter Notebooks
MLFlow • Fairlearn • AIF360
Section 15.0
API Connect
Secured APIs
Enabler Tools (IBM-Managed)
Data Governance & Quality
• IBM Knowledge Catalog (IKC)
• IBM Manta Data Lineage
Data Intelligence
• watsonx.data Intelligence
(Ingestion & Cataloging)
AI Trust & Governance
• watsonx.governance
(Bias Detection, Fairness)
Commons Credits System
PostgreSQL • Orchestrate • Apptio
(Sections 7.0-7.7)
Synthetic Data Generation
• watsonx.ai (Tabular & Text)
Security Gateway
• IBM API Connect
(OAuth 2.0, RBAC, Rate Limits)
Sections 1.0, 5.0, 8.0-10.0, 13.0
Snowflake Data Warehouse
• Core Data Processing & Storage
• Scalable CPU/GPU Compute (SPCS)
• QUERY_HISTORY & ACCESS_HISTORY
• Data Versioning & Encryption
• RBAC & Tamper-Evident Logs
Sections 5.0, 7.1, 8.0, 20.0
MGHPCC Data Center
• AI Compute Resources (HPC)
• Dataset Hosting (Hybrid)
Table 1 (Optional)
Ingestion
Metadata
Data Sources
• External Sources (Hugging Face, World Bank)
• On-Premises & Cloud Storage
• Structured, Semi-Structured, Unstructured Data
Section 5.0
Key Integration Points
Security: IBM API Connect secures all interactions (OAuth 2.0, RBAC, MFA) - Section 1.0, 19.0
Compliance: FAIR Principles, HIPAA, EU AI Act alignment - Sections 1.0, 14.0, 20.0
Scalability: Snowflake provides elastic compute and storage - Section 8.0
Collaboration: IBM Expert Labs + SI + Snowflake iterative refinement - Section 3.0

Figure 1. MassTech AI Hub Data Commons Collaborative (DCC) Technical Architecture

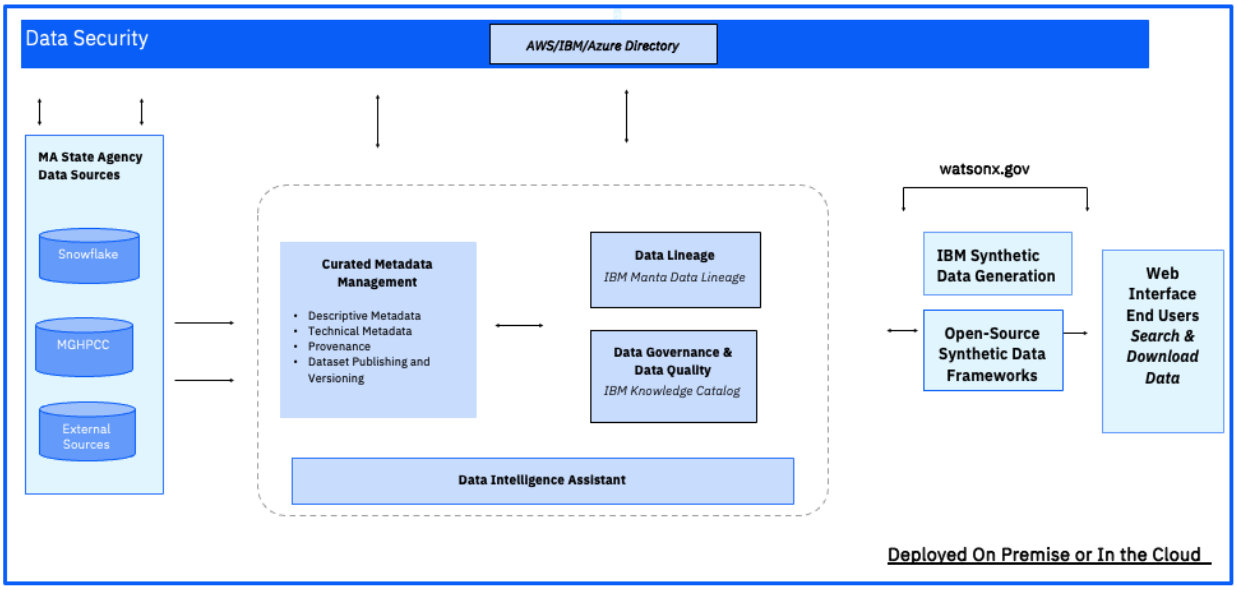
2.2 IBM Technology Components

2.2.1 Data Catalog and Metadata Management System

A data catalog and metadata management system needs to be implemented to manage hosted data that is curated and loaded to the data platform. This section represents the requirements for this capability. The solution shall organize, describe, and make datasets discoverable and reusable by managing their associated metadata. It enables users to publish, search, and access data with clearly defined context—such as who created it, what it contains, how it can be used, and where it came from. This is especially useful in a data commons because it supports FAIR data principles (Findable, Accessible, Interoperable, Reusable), facilitates collaboration and governance, and ensures that data is properly documented, preserved, and cited over time.

As exemplified in the diagram below IBM watsonx.data integrates with both Snowflake and on-premises sources by syncing metadata through secure connectors, delivering a unified catalog that captures descriptive (titles, authors, keywords), technical (format, schema, size, dates), and provenance metadata (origin, transformations) using open standards like Dublin Core, Schema.org, and DCAT. The solution enables faceted and full-text search across all assets, plus programmatic access via APIs, allowing users to discover, filter, and govern datasets in hybrid environments with strong lineage and

compliance support.



Metadata Management Components

IBM Knowledge Catalog (IKC) in Watsonx.Data Intelligence is a robust, enterprise-grade metadata catalog that addresses the requirements of the DCC. This solution can organize, describe and make datasets discoverable and reusable by managing the associated metadata. It will track and manage the metadata of data assets that are uploaded into the DCC, allowing for clear context like who created the asset and where it came from, what the asset contains and how it can be used.
Descriptive metadata like title, author, abstract, keywords, and data use/license can be captured and saved in the metadata of an asset through the IBM Knowledge Catalog (IKC) service in Watsonx.Data Intelligence, as long as the asset includes these values. When the metadata is scanned and imported, the system will capture these values. Then a set of business terms, data classes and classifications can be assigned to them in order to standardize across data sets and provide searchability and rules/policies across these values. For any values not picked up in the metadata, custom terms/tags can be applied to the data to capture these values. Technical metadata like file format, schema, size, date of creation/update are also all automatically captured as standard metadata of assets ingested into the data catalog.

The provenance of data assets is tracked when ingested into the data catalog by automatically noting the data source and path, and the user that uploaded it. This information is available at the asset level along with the general asset metadata. Extensive lineage of datasets is also tracked via the industry-leading data lineage solution, Manta, that tracks system hops and transformations from source to target. This gives a DCC admin, and possibly DCC users, the ability to learn where data came from, how it was transformed, and where the data was moved. This lineage data is displayed to the admin via a graphical interface, visually showing data lineage and transformations. This can be useful for data incident resolution, impact analysis, regulatory compliance, and improving trust in data.

This data catalog can adhere to the metadata standards of Dublin Core, Schema.org, or DCAT. It is built with its own metadata model and governance framework. Because it is designed for interoperability and extensibility- you can incorporate or align with external standards by customizing your business glossary and metadata model. IBM's data catalog is a proven enterprise ready solution indicated by IDC's latest

data intelligence magic quadrant.

Dataset Discovery and Search

IBM Knowledge Catalog provides comprehensive backend catalog capabilities enabling global dataset discovery through keyword searches and advanced filtering options including tags, creators, dates, and file formats. Users can execute sophisticated queries using key-value pairs to search across asset properties such as descriptions, custom metadata fields, tags, and column names. The catalog supports full-text search across all metadata elements, focusing on structured metadata rather than file content to ensure efficient, scalable performance.

The search engine delivers two complementary result types optimized for different discovery patterns. AI-powered semantic search leverages large language models to return contextually relevant results including synonyms and related concepts, clearly marked with AI indicators in the catalog interface. Lexical search provides exact-match results from metadata properties and column names, ensuring precise retrieval when users know specific terminology. This dual-mode approach accommodates both exploratory research workflows and targeted data retrieval needs across MassTech's diverse user communities.

All search functionality is fully accessible through IBM Knowledge Catalog's comprehensive API suite, enabling the System Integrator to build intuitive, user-facing search interfaces within the Data Commons portal. These RESTful APIs support programmatic search and retrieval operations, allowing the SI to:

- Execute detailed asset searches using backend metadata without exposing catalog complexity to end users
- Retrieve complete metadata for selected datasets to populate rich detail pages
- Generate previews of structured data assets within frontend interfaces
- Manage asset lifecycle operations including publication, updates, and deprecation through

administrative APIs

This API-first architecture provides MassTech with maximum flexibility to customize catalog functionality exposed to researchers, government agencies, and private sector innovators as the Data Commons platform evolves. The separation between backend catalog services and frontend user experience enables continuous refinement of discovery workflows based on user feedback while maintaining enterprise-grade governance and performance standards throughout the metadata management layer.

2.2.2 Synthetic Data Tools

IBM supports the Data Commons' commitment to open-source innovation while recommending the inclusion of IBM watsonx.ai synthetic data generators to enhance the platform's data generation capabilities. These enterprise-grade tools complement open-source alternatives by providing specialized functionality for privacy-preserving analytics and AI model development across diverse research use cases.

Synthetic Data Generation Capabilities

IBM watsonx.ai delivers two complementary synthetic data generation tools addressing distinct research requirements:

Tabular Data Generator: A visual flow editor enables researchers to create or replicate structured datasets mimicking real-world data distributions without exposing sensitive information. This capability supports AI model training, testing scenarios, and privacy-preserving analytics where actual datasets cannot be shared due to regulatory constraints, confidentiality agreements, or ethical considerations. Researchers can define schemas, specify statistical properties, and generate realistic synthetic datasets that preserve analytical utility while eliminating disclosure risks.

Unstructured Text Generator: Foundation model-powered text synthesis produces contextually relevant synthetic text for specialized applications including tool calling, text-to-SQL translation, and domain-specific question answering. This capability enables researchers to augment limited training datasets, test natural language processing models, and develop AI assistants without requiring access to proprietary or sensitive textual content. The beta designation reflects ongoing enhancements incorporating latest advances in large language model capabilities.

Platform Integration and Accessibility

Both generators are accessible through watsonx.ai Studio for interactive development and watsonx.ai Runtime for programmatic generation at scale. IBM Knowledge Catalog automatically captures and surfaces comprehensive metadata for each synthetic dataset including generation methodology, source parameters, applicable licenses, and usage guidelines. The System Integrator's frontend interface can render tool descriptions, access links, and illustrative use case examples helping researchers select appropriate generation approaches for their specific requirements.

API-Driven Architecture

All synthetic data generation functionality is fully accessible via RESTful APIs enabling seamless integration within the Data Commons' open-source infrastructure. The SI can incorporate generation capabilities directly into research workflows, allowing users to:

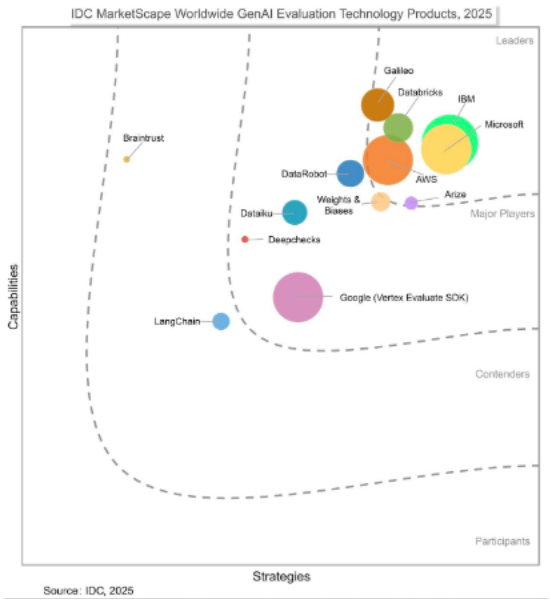
- Initiate synthetic data generation jobs programmatically from within the portal interface
- Monitor generation progress and retrieve completed datasets through standardized API endpoints
- Query available generation templates and parameter options to guide user selections
- Integrate synthetic datasets with existing analysis pipelines and computational resources

This API-first design ensures watsonx.ai synthetic data generators interoperate with open-source tools and frameworks adopted by MassTech, providing researchers with flexible options combining IBM's enterprise capabilities with community-driven innovation. The architecture supports MassTech's goal of building an extensible platform where commercial and open-source components collaborate seamlessly to accelerate responsible AI research across the Commonwealth.

2.2.3 AI Fairness and Bias Capability

IBM AI Governance Overview

Watsonx.governance (watsonx.gov) supports the AI Fairness and Bias Capability within the Data Commons by providing a comprehensive framework to ensure that data and AI systems are equitable, transparent, and aligned with ethical standards. It enables the identification of bias in datasets, models, and outcomes through profiling and fairness testing across sensitive attributes. To mitigate bias, it offers pre-processing, in-processing, and post-processing techniques that can be integrated into model development workflows. Watsonx.gov also facilitates ongoing monitoring of deployed models to detect fairness drift and trigger alerts when thresholds are breached, ensuring accountability over time. Additionally, it promotes transparency through model documentation, policy enforcement, and stakeholder collaboration, making it well-suited for high-impact public sector applications where ethical alignment and compliance are critical. As evidenced in the latest IDC report, watsonx.gov is the industry leader in AI evaluation technology enabling Mass Tech to automate evaluation today and for the future.



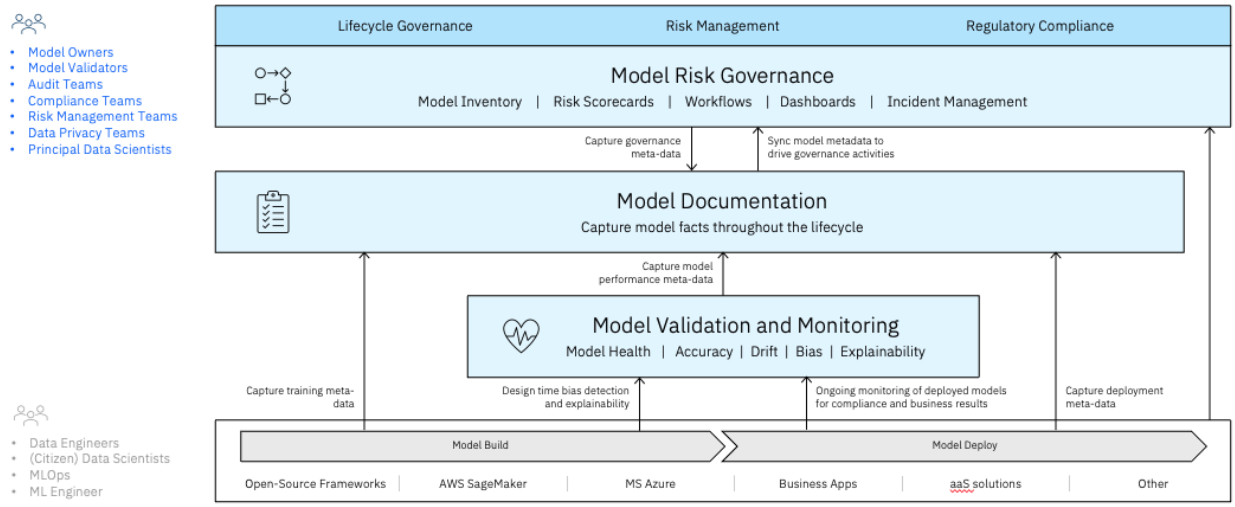
Human Centric, Trustworthy, Inclusive, Innovative, & Sustainable

Watsonx.gov aligns closely with the principles outlined for ethical AI and data governance. It is designed to enhance human autonomy by supporting responsible decision-making rather than replacing it, ensuring that AI systems contribute positively to societal wellbeing. The platform fosters trust through robust features such as automated model documentation, lineage tracking, and risk monitoring, which collectively promote security, reliability, explainability, and transparency. These capabilities help uphold ethical norms and legal compliance, reinforcing public confidence in AI systems. Additionally, watsonx.gov actively addresses bias and promotes equity by enabling bias detection and mitigation, encouraging diverse data representation, and supporting inclusive model development. Its modular architecture and integration capabilities facilitate open collaboration and rapid experimentation, making it a powerful tool for ethical innovation and cross-sector partnerships. Finally, IBM's commitment to sustainability is reflected in the platform's support for resource efficient model management and environmentally conscious infrastructure, ensuring long-term viability and alignment with intergenerational justice.

Legal and Regulatory Guidance

IBM watsonx.governance addresses legal and regulatory guidance requirements by embedding compliance, transparency, and ethical oversight into its governance workflows that tie to an AI Use with automated AI lifecycle metadata collection and metric monitoring capabilities.
Watsonx.gov helps organizations avoid unauthorized commercialization of AI models and fair use compliance by enforcing strict governance across the entire AI lifecycle. It ensures that every model is tied to an approved use case through structured workflows and risk assessments, preventing deployment outside authorized contexts. AI factsheets and centralized model inventories provide transparency and traceability, making it easy to detect misuse. By translating regulatory requirements into enforceable internal policies and offering explainability tools, Watsonx.gov ensures that AI models are used ethically, legally, and only within approved commercial boundaries. To meet anti-kickback statute requirements, the platform includes a regulatory compliance solution that tracks relevant laws, links data-sharing use cases to regulatory mandates, and updates workflows as regulations evolve. It also supports risk and applicability assessments to proactively identify legal and ethical risks, ensuring that data-sharing agreements do not resemble improper financial incentives. Automated workflows and

audit trails provide transparency and accountability in decision-making, while metadata monitoring ensures data is used fairly and not manipulated for gain. In terms of nondiscrimination policies, watsonx.gov offers robust bias detection and mitigation tools, including a Bias Scanner and Intersectional Fairness Analysis, which evaluate models across demographic variables and combinations to prevent disparate treatment. Continuous monitoring helps maintain fairness over time. The platform also enables policy enforcement through customizable workflows, role-based access controls, and audit trails, ensuring that nondiscrimination policies are actively upheld. Its ethical AI governance framework emphasizes fairness, transparency, and human oversight, aligning with global standards such as the EU AI Act and NIST AI RMF. Finally, watsonx.gov supports the structuring of data-sharing agreements under a "Data as a Service" model by embedding compliance checks, linking use cases to regulations, automating risk assessments, and maintaining audit trails. These capabilities collectively ensure that organizations can meet legal and regulatory obligations while promoting ethical and fair AI practices.



Ethical Considerations

IBM watsonx.governance embeds ethical principles by offering standardized tools and workflows that address key ethical requirements. It provides mechanisms for quantifying and reducing algorithmic bias through its Bias Scanner, which evaluates models using over 120 demographic variables and 18 fairness metrics, and integrates with CI/CD pipelines for automated, auditable testing. The platform supports regular fairness assessments, especially in high-risk domains like healthcare and finance, through continuous monitoring dashboards, intersectional fairness analysis, and automated governance workflows. For privacy risk mitigation, watsonx.governance employs zero data retention, encryption, secure deletion protocols, metadata management, and PII scanning, aligning with techniques such as differential privacy, k-anonymity, and secure multi-party computation. It also promotes transparency in AI decision-making via model explainability tools, audit trails, and real-time monitoring, while enabling organizations to meet regulatory standards that support access to legal remedies. Additionally, it helps guard against identification risks when external data is combined with DCC datasets by managing metadata and quasi-identifiers, ensuring ethical compliance throughout the AI lifecycle. It does this by managing metadata and quasi-identifiers through a robust, policy-driven governance framework that spans the entire AI lifecycle. At the core of this system are AI Factsheets, which automatically capture and document metadata such as data lineage, model provenance, and usage context. This metadata helps identify quasi-identifiers— attributes like ZIP code or birthdate that may not be directly identifying

on their own but can become sensitive when combined with other data.

2.2.4 IBM API Connect

IBM API Connect serves as the secure and scalable API management layer within the Massachusetts AI Hub Data Commons Collaborative, functioning as the strategic enterprise asset that enables integration across diverse data sources while supporting digital transformation in AI research and governance. As the unified gateway orchestrating interactions between the System Integrator's web portal, IBM's governance and credit systems, and Snowflake's data platform, API Connect ensures the DCC's curated datasets and synthetic data capabilities are accessible in a secure, governed, and scalable manner aligned with MassTech's vision for transparency, equitable access, and cross-sector collaboration.

Full API Lifecycle Management

IBM API Connect covers the complete API lifecycle—from creation and security through management and analytics—ensuring that curated data services within the Data Commons can be exposed, governed, and scaled securely throughout their operational lifespan. This comprehensive lifecycle management supports the DCC's evolution as new datasets, credit systems, and synthetic tools are onboarded, with version control and deprecation policies that maintain backward compatibility while enabling continuous innovation. The platform's cloud-native, microservices-based deployment architecture supports extensibility and hybrid models, directly addressing the DCC's future-proofing goals and enabling seamless integration across Massachusetts datasets, global repositories, and synthetic data generation tools.

Enterprise-Grade Security and Governance

The DCC's requirements for strict governance and compliance—including role-based access controls, comprehensive audit trails, and transparency—are directly addressed through API Connect's enforcement of OAuth 2.0, OpenID Connect, API keys, and threat protection policies. The platform implements a trusted, security-hardened gateway with an extensive range of policies protecting APIs while supporting enterprise standards and security protocols including WS-Security, SAML, and Kerberos. Integrated API Governance enforces organizational standards for consistency and quality across all data services, while API Discovery capabilities identify shadow and unmanaged APIs running within the estate to ensure comprehensive oversight. Advanced security integration with Akamai enables real-time detection and prevention of malicious attacks, complementing Snowflake's encryption at rest and in transit to create defense-in-depth protection for sensitive datasets.

Multi-Form API Management for Diverse Data Services

API Connect provides first-class support for diverse API formats including REST, AsyncAPIs, GraphQL, SOAP, and WebSocket, enabling the DCC to expose varied data services through unified management infrastructure. Declarative GraphQL creation from federated data sources using IBM API Connect for GraphQL allows complex queries across multiple datasets stored in Snowflake, while AsyncAPI management through integration with Event Endpoint Management supports real-time data streaming for dynamic AI applications. This capability ensures that all API types—whether accessing static datasets, streaming synthetic data generation, or querying lineage information via Manta—can be shared on the same developer portal for easy discovery and consumption, reducing fragmentation and improving the researcher experience.

Hybrid and Open Architecture

API Connect's support for multi-cloud deployment provides flexible hybrid cloud capabilities to deploy different components on any cloud platform or on-premises infrastructure, with native Kubernetes-based deployment supporting IKS, AWS EKS, Azure AKS, Red Hat OpenShift (ROKS, ROSA, ARO), and Open Virtual Appliance configurations. Gateway choice and flexibility accommodate all use cases—internal services, secure external networks, and traditional applications—ensuring that whether datasets originate from university repositories, state government systems, or commercial providers, they flow through consistent, governed pathways. This architectural flexibility directly supports the DCC's goal of connecting public, private, and academic sectors while maintaining appropriate security boundaries and compliance requirements for each stakeholder community.

AI-Enhanced API Operations

The purpose-built AI Gateway within API Connect provides policies specifically designed for AI services, including token-based cost management that integrates with the Commons Credits System, response caching to optimize performance for frequently accessed datasets, and comprehensive insights into the organization's use of AI services that enable business leaders to make data-driven decisions about AI service adoption and resource allocation. Generative AI-powered API Assistant capabilities accelerate productivity across API development teams by enhancing code quality, improving API documentation, and upskilling developers for faster time-to-market—particularly valuable for academic researchers and startups who may lack extensive API expertise but need to quickly integrate DCC capabilities into their applications.

Developer Enablement and Community Growth

The IBM API Connect Developer Portal provides a branded, self-service interface where researchers, startups, and government agencies can discover and subscribe to available data APIs, register applications, manage credentials, and access comprehensive documentation with usage analytics and onboarding resources. This portal reduces barriers to adoption by empowering users to independently explore the data ecosystem, understand API capabilities through interactive examples, and track their consumption against credit balances in real-time. The portal supports community growth by enabling data contributors to publish APIs for their datasets—cataloged through IBM Knowledge Catalog and stored in Snowflake—while consumers explore and integrate services that accelerate AI innovation, fostering the cross-sector collaboration central to MassTech's vision.

Integration with Commons Credits System

API Connect enforces rate limits and quotas aligned with the Commons Credits System's debit model, with consumers debited based on API usage metrics including call volume and data transfer. Rate plans are customizable per user, organization, or API product to reflect tiered consumption levels, while usage metrics are exported to IBM Apptio for integration with the credit engine, enabling real-time correlation between API activity and transactions stored in the PostgreSQL immutable ledger. Dataset uploads trigger automated credit allocation workflows, with API Connect assessing metadata quality and demand signals to determine contribution value. The /api/v1/credits/balance/{user_id} endpoint provides real-time balance and transaction history, while vendor cost pass-through for third-party services and peer-to-peer transactions are orchestrated through the integration hub with API Connect managing secure authentication and usage tracking.

2025 IBM CORPORATION | CLIENT AND IBM CONFIDENTIAL

Monitoring, Governance, and Transparency

Built-in analytics provide comprehensive visibility into API usage patterns, consumer behavior, traffic volumes, and policy enforcement metrics, with logs forwarded to observability platforms or SIEM tools for compliance auditing and operational insights. Immutable audit trails track all API interactions as tracked, auditable interfaces supporting the DCC's transparency requirements, while real-time public dashboards display system health and credit circulation. Open APIs enable third-party verification of platform operations, with public documentation of all policies and governance decisions accessible through the Developer Portal. This transparency infrastructure ensures APIs enforce role-based permissions aligned with the DCC's fairness and ethical use mandates, creating equitable access pathways for all stakeholder communities.

API-Driven Governance and Metadata Management

API Connect powers the integration of IBM Knowledge Catalog (IKC) and IBM Manta Data Lineage with the SI's web portal through secure, RESTful endpoints that follow consistent design principles:

- **/api/v1/catalog/search**: Full-text and faceted search across IKC enabling dataset directory discovery
- **/api/v1/assets/{id}/metadata**: Detailed metadata retrieval for rich dataset detail pages
- **/api/v1/lineage/{asset_id}**: Graphical data provenance via Manta supporting compliance verification
- **/api/v1/credits/balance/{user_id}**: Real-time credit balance and transaction history
- **/api/v1/governance/bias/scan**: Fairness evaluation initiation through watsonx.governance

Comprehensive OpenAPI specifications enable automatic client generation, integration guides support rapid onboarding, and full-featured sandbox environments allow testing without impacting production systems. IBM Expert Labs engineers maintain dedicated support channels for troubleshooting, while the platform's event-driven architecture and service mesh capabilities ensure real-time processing with advanced networking and observability.

Strategic Long-Term Transformation

IBM positions API Connect as the strategic on-ramp to digital innovation for the Data Commons, embedding APIs into the platform's fabric to support scaling and future expansion. This strategic role encompasses enabling integration across diverse data sources (Massachusetts datasets, global repositories, synthetic tools), supporting digital transformation in AI research and governance by making data services programmatically accessible, and providing a foundation where new data services and tools can be onboarded seamlessly as the ecosystem evolves. By centralizing API access and management within this enterprise-grade platform, the DCC ensures its curated datasets and synthetic data capabilities remain accessible, governed, and scalable as Massachusetts solidifies its position as a national leader in responsible AI innovation.

2.2.5 Commons Credits System

IBM proposes implementing the Commons Credits System in a two-phase approach to ensure a secure, immutable ledger foundation. It is governed by clear workflows and transaction rate limits and seamlessly integrated with Snowflake’s data platform and the System Integrator’s (SI) user-facing applications. IBM

API Connect plays a key role in securing and coordinating all interactions, enabling controlled access and scalability while supporting MassTech's goals for efficient AI resource management.

Phase 1: Foundation, Workflows, and Rate Setting (SI Timeline Integration - TBD)

IBM PostgreSQL acts as a secure, auditable ledger for the Commons Credits System, maintaining an unalterable history of all credit transactions:

- **Audit Tables for Immutability and History**: A PostgreSQL schema is designed with primary data tables for current credit balances and separate Audit Tables that mirror them for traceability.
- **Trigger Implementation**: Database triggers automatically record every change into the corresponding Audit Tables, creating tamper-proof audit trails.
- **Raw Data Integration**: Secure connectors managed via IBM API Connect extract raw usage logs from Snowflake's QUERY_HISTORY and ACCESS_HISTORY views, enabling precise credit calculations based on data processing and storage activities.

Initial logic and rate setting are implemented using IBM watsonx Orchestrate and IBM API Connect:

- **Transaction Workflows**: IBM watsonx Orchestrate manages credit debit and credit transactions, integrating with Snowflake to automate credits based on usage metrics.
- **Security and Rate Setting with IBM API Connect**: IBM API Connect enforces OAuth 2.0 authentication, rate limits, and secure API endpoints (/api/v1/credits/balance/{user_id} for real-time balance and transaction history), ensuring controlled, secure access between the SI's web portal and the backend ledger. API Connect's centralized gateway enforces rate-limiting policies and monitoring to prevent overuse.

Data Commons Collaborative Ledger Tables

List of Ledger Tables

Table Name	Purpose / Notes	Reference Documentation
DCC_Ledger_INTERFACE	Main open interface table for credit transaction imports. Rows inserted here are validated and converted to credit journal entries for the Commons Credits System. Serves as the primary ingestion point for credit debits (resource consumption) and credits (dataset contributions).	IBM PostgreSQL Documentation
DCC_Ledger_INTERFACE_CONTROL	Controls the execution of Credit Transaction Import processes including validation status, transaction source identification (Snowflake usage metrics, API Connect activity, manual adjustments), and unique run identifiers for batch processing and audit trail correlation.	IBM PostgreSQL Documentation
DCC_Ledger_SEGMENT_VALUES_INTERFACE	Interface table used to load segment values for credit allocation hierarchies including user organizations, dataset categories, resource types (storage, compute, API calls), and quality tier classifications supporting tiered crediting structures.	IBM watsonx Orchestrate Documentation
DCC_Ledger_SEGMENT_HIER_INTERFACE	Interface table to load segment value hierarchies defining organizational structures, dataset taxonomies, and resource allocation categories. Supports nested credit allocation models enabling	IBM watsonx Orchestrate

	enterprise-wide, departmental, and project-level credit tracking with rollup capabilities.	Documentation
DCC_Ledger_BALANCES	Primary table storing current credit balances for all users, organizations, and projects. Optimized for high-frequency read operations supporting real-time balance queries via /api/v1/credits/balance/{user_id} API endpoint. Updated through triggers from transaction tables.	IBM API Connect Documentation
DCC_Ledger_TRANSACTIONS	Core transaction table recording all credit movements including debits (resource consumption, third-party service costs) and credits (dataset contributions, quality bonuses). Each row represents an immutable transaction with timestamp, source reference, and correlation to Snowflake usage logs.	IBM PostgreSQL Documentation
DCC_Ledger_AUDIT	Audit table mirroring all changes to balances and transactions, creating tamper-proof audit trails through database triggers. Captures complete transaction history with before/after values, user identifiers, timestamps, and source system references supporting transparency and third-party verification requirements.	IBM PostgreSQL Documentation
DCC_Ledger_RATES	Configuration table defining credit earning rates (contribution valuation, metadata quality multipliers) and spending rates (compute resource conversion factors, storage pricing, API call costs). Supports tiered rate structures and dynamic pricing adjustments based on platform utilization and demand elasticity.	IBM Apptio Documentation
DCC_Ledger_SOURCES	Reference table cataloging all transaction sources including Snowflake QUERY_HISTORY extracts, watsonx.data Intelligence quality scores, API Connect usage metrics, manual administrator adjustments, and third-party AI service billing. Enables source-based reporting and reconciliation.	IBM API Connect Documentation
DCC_Ledger_RECONCILIATION	Table supporting periodic reconciliation processes correlating PostgreSQL ledger transactions with Snowflake audit logs, API Connect usage analytics, and Apptio financial reports. Identifies discrepancies, validates data integrity, and generates compliance attestations for transparency requirements.	IBM Apptio Documentation

Additional Ledger Support Tables

Table Name	Purpose / Notes	Reference Documentation
DCC_Ledger_USERS	User registry table storing unique identifiers, organizational affiliations, authorization levels (researcher, administrator, auditor), and credit allocation policies. Integrates with IBM API Connect OAuth 2.0 authentication and role-based access control enforcement.	IBM API Connect Documentation
DCC_Ledger_ORGANIZATIONS	Organization hierarchy table defining academic institutions, government agencies, and private sector entities participating in the Data Commons. Supports organizational credit pools, shared resource allocations, and departmental rollup reporting through nested structures.	IBM watsonx Orchestrate Documentation
DCC_Ledger_DATASETS	Dataset registry table linking dataset identifiers from IBM Knowledge Catalog to credit earning events. Tracks contribution dates, quality scores, download frequency, and citation metrics influencing credit allocation calculations and premium rate eligibility.	IBM Knowledge Catalog Documentation
DCC_Ledger_RESOURCES	Resource definition table cataloging billable platform resources including Snowflake compute warehouses (small, medium, large, x-large), storage tiers (standard, archive), API endpoints, and third-party AI services. Maps resources to credit spending rates and usage metrics extraction patterns.	Snowflake Documentation
DCC_Ledger_WORKFLOWS	Workflow configuration table defining automated credit allocation rules managed by IBM watsonx Orchestrate. Includes validation logic (balance sufficiency checks), approval routing (high-value transactions), notification triggers (low balance alerts), and scheduled batch processes (nightly reconciliation).	IBM watsonx Orchestrate

DCC_Ledger_EXCEPTIONS	Exception tracking table capturing failed validations, rejected transactions, and data quality issues requiring manual intervention. Supports exception workflow routing to administrators, resolution tracking, and root cause analysis improving system reliability.	Documentation IBM watsonx Orchestrate Documentation
DCC_Ledger_METRICS	Performance metrics table recording transaction processing latency, API response times, reconciliation completion durations, and system health indicators. Feeds operational dashboards in IBM Apptio and supports capacity planning and performance optimization initiatives.	IBM Apptio Documentation

Table Relationship Architecture

The DCC_Ledger table structure implements a comprehensive audit and governance framework:

1. **Transaction Flow:** DCC_Ledger_INTERFACE → DCC_Ledger_INTERFACE_CONTROL → DCC_Ledger_TRANSACTIONS → DCC_Ledger_BALANCES
2. **Audit Trail:** DCC_Ledger_AUDIT mirrors all changes to DCC_Ledger_TRANSACTIONS and DCC_Ledger_BALANCES through database triggers
3. **Reference Data:** DCC_Ledger_RATES, DCC_Ledger_SOURCES, DCC_Ledger_RESOURCES provide configuration supporting transaction validation
4. **Hierarchical Structures:** DCC_Ledger_SEGMENT_VALUES_INTERFACE and DCC_Ledger_SEGMENT_HIER_INTERFACE define organizational and resource allocation hierarchies
5. **Reconciliation:** DCC_Ledger_RECONCILIATION correlates PostgreSQL ledger with external systems (Snowflake, API Connect, Apptio) ensuring data integrity

This architecture ensures the Commons Credits System maintains immutable transaction history, supports real-time balance queries with sub-second latency, enables transparent audit trails for compliance verification, and provides foundation for sophisticated financial intelligence through IBM Apptio analytics.

Phase 2: Analytics, Optimization, and Advanced Governance (SI Timeline Integration - TBD)

- **Earning Commons Credits**: IBM watsonx Orchestrate automates workflows to issue credits based on metadata quality, validated through Snowflake's data processing capabilities and accessible via secure API Connect endpoints.
- **Spending Commons Credits**: IBM Apptio manages spending policies through API Connect endpoints, tracking consumption against Snowflake's compute and storage resources.
- **Incentivizing Efficient Usage**: IBM Apptio monitors resource metrics for tiered crediting, utilizing Snowflake's scalable infrastructure to optimize credit distribution based on usage efficiency.
- **Governance and Transparency**: IBM Apptio dashboards, integrated via API Connect, link transactions to immutable logs stored in PostgreSQL, with Snowflake ensuring end-to-end auditability. API Connect provides audit logging for all API interactions, promoting transparency.
- **User Experience & Access Control**: IBM Apptio visualizations are incorporated into the SI's DCC portal via API Connect, supporting role-based access control (RBAC) aligned with Snowflake's secure data environment.
- **Technical Expectations**: RBAC, multi-factor authentication (MFA), and encryption are implemented with audit trails, managed through API Connect's security policies and Snowflake's compliance features, ensuring adherence to standards.

This system enhances the SI's front-end development by offering robust, API-driven backend functionality through IBM API Connect. It guarantees scalability, security, and transparency. IBM will work with the SI to define and test API specifications, leveraging Snowflake for data handling and API

Connect for secure orchestration to deliver a cohesive, user-focused experience aligned with MassTech's vision for the Data Commons Collaborative.

2025 IBM CORPORATION | CLIENT AND IBM CONFIDENTIAL

3. IBM Expert Labs Implementation Services

Integrated Solution Delivery Through Strategic Partnership

IBM's Expert Labs collaboration with the selected System Integrator (SI) and Snowflake will address requirements. The SI will confirm requirements and scope, objectives and approach. The scope, objectives and approach will be revised to integrate to the Program office as well as relevant project plans.

Overview

IBM Expert Labs delivers comprehensive implementation services for the Massachusetts AI Hub Data Commons Collaborative through a structured, phased approach that integrates specialized expertise in data governance, metadata management, API orchestration, and financial intelligence. Our implementation strategy leverages proven methodologies from IBM Technology Expert Labs service offerings—including Build Data Catalog (SaaS), Build Data Quality (SaaS), Manta Data Lineage QuickStart SaaS, Data Platform Accelerator Services, and Architect MQ for IBM API Connect—combined with Apptio Jumpstart services for the Commons Credits System. This consolidated implementation establishes a secure, interoperable, and scalable data platform aligned with MassTech's vision for responsible AI innovation across the Commonwealth.

Phased Implementation Approach

Phase 1: Foundation and Governance

IBM Expert Labs establishes the foundational infrastructure for data governance, API orchestration, and credit system architecture during the initial implementation phase, creating the backbone for secure, auditable data sharing and resource management.

Data Catalog Foundation

IBM deploys watsonx.data Intelligence with a cross-industry Knowledge Accelerator through an 80-hour consulting engagement that establishes comprehensive metadata management capabilities. The implementation configures Data Asset Catalogs enabling global search and relationship exploration, with metadata import and enrichment processes ingesting initial datasets from MassTech's priority data sources. IBM Expert Labs implements up to five Data Protection Rules enforcing role-based access controls and audit trail requirements, ensuring that researchers, government agencies, and private sector innovators access data appropriate to their authorization levels. Integration with IBM API Connect enables the System Integrator to build user-facing search interfaces consuming /api/v1/catalog/search and /api/v1/assets/{id}/metadata endpoints, providing intuitive dataset discovery aligned with FAIR principles.

API Orchestration Blueprint

A dedicated IBM architect leads discovery sessions with MassTech stakeholders through whiteboard workshops, staff interviews, and existing system assessments to create a comprehensive API Connect

deployment blueprint. The architect analyzes infrastructure requirements across hybrid cloud environments, designs high availability and disaster recovery configurations ensuring continuous platform operation, and defines security architectures implementing OAuth 2.0, OpenID Connect, and threat protection policies. The blueprint documents configuration and topology recommendations optimizing message flow between the SI's web portal, IBM governance components, and Snowflake's data platform, while establishing DevOps and monitoring strategies that integrate with MassTech's existing CI/CD pipelines. This deliverable provides the technical foundation for all API-driven interactions across the Data Commons ecosystem.

Commons Credits System - Foundational Architecture

The Commons Credits System implementation leverages Apptio Jumpstart for Hybrid Business Management Infrastructure Costing to establish cost transparency and allocation capabilities essential to credit tracking. IBM Expert Labs facilitates business requirements workshops defining credit earning rules—including dataset contribution valuation, metadata quality thresholds, and data sharing incentives—alongside spending policies governing compute resource consumption, premium data access tiers, and third-party AI service usage. The team conducts current state assessments documenting existing Snowflake usage patterns through QUERY_HISTORY and ACCESS_HISTORY analysis, identifying baseline consumption metrics and cost allocation opportunities.

IBM designs the technical architecture integrating multiple components: IBM PostgreSQL serves as the immutable ledger with primary tables storing current credit balances and separate audit tables capturing complete transaction history through database triggers; IBM watsonx Orchestrate manages automated workflows for credit debits (resource consumption) and credits (contributions), with validation logic ensuring balance sufficiency before resource allocation; IBM API Connect enforces rate limiting, secure authentication via OAuth 2.0, and comprehensive audit logging for the /api/v1/credits/balance/{user_id} endpoint providing real-time balance queries.

Apptio platform configuration establishes cost model templates adapted for credit allocation, with hierarchies mapping user organizations, dataset categories, and resource types including storage, compute, and API calls. IBM implements secure connectors extracting Snowflake usage metrics and watsonx.data Intelligence quality scores, feeding automated data pipelines that correlate platform activity with credit transactions. User acceptance testing validates credit earning for sample dataset uploads, spending workflows for compute requests, and audit trail integrity before production deployment. The phase concludes with administrator training on dashboard usage, policy management, and ledger maintenance procedures.

Phase 1 Deliverables:

- Operational IBM Knowledge Catalog with initial datasets cataloged and searchable via API
- IBM API Connect implementation blueprint documenting infrastructure specifications, security architecture, high availability design, and phased deployment roadmap
- Functional Commons Credits System with PostgreSQL ledger, watsonx Orchestrate workflows, and Apptio cost model configured for credit allocation
- Integration of Snowflake usage data feeding credit calculations and audit trails
- Administrator training materials and operational documentation for foundational components

Phase 2: Advanced Capabilities and Analytics

IBM Expert Labs expands platform capabilities during Phase 2, implementing advanced data quality management, comprehensive lineage tracking, and sophisticated financial intelligence that transform the Data Commons from a foundational infrastructure into a fully-featured innovation hub.

Data Quality Framework Implementation

IBM deploys comprehensive data quality capabilities through an 80-hour consulting engagement establishing trust in AI-ready datasets. The implementation configures five Data Quality Definitions encompassing completeness, accuracy, consistency, timeliness, and validity dimensions, with up to 20 specific rules monitoring critical data attributes across the catalog. IBM establishes SLA workflows automating quality assessments at scheduled intervals and triggering remediation alerts when datasets fall below defined thresholds, ensuring that researchers receive only trustworthy data for AI model development.

Integration with the Data Catalog ensures quality metrics appear prominently in dataset metadata, influencing credit allocation through quality score multipliers that reward high-quality contributions with premium credit rates. IBM executes metadata import and enrichment jobs for connected data sources, establishing baseline quality scores and configuring automated profiling that continuously monitors data integrity as new datasets enter the platform. The SI leverages these quality indicators through API Connect endpoints, displaying trust scores and quality certifications that help users make informed decisions about dataset selection for their research initiatives.

Data Lineage and Provenance Tracking

IBM implements Manta Data Lineage through a 104-hour fixed-scope engagement providing end-to-end traceability essential for regulatory compliance and trust-building. The implementation begins with Manta software installation on Cloud Pak for Data or compatible infrastructure, followed by configuration of up to five supported scanner connections including one ETL scanner capable of ingesting maximum 20,000 scripts. IBM conducts discovery workshops with MassTech subject matter experts to identify priority data sources requiring lineage tracking, focusing on datasets with complex transformation pipelines or stringent compliance requirements.

Lineage Reviews validate data flow accuracy for three initial data sources, documenting how external datasets from sources like Hugging Face, World Bank, or Massachusetts state agencies flow through Snowflake transformations before appearing in the catalog. IBM integrates lineage metadata into IBM Knowledge Catalog, enabling graphical visualizations accessible via the /api/v1/lineage/{asset_id} API endpoint that the SI incorporates into dataset detail pages. These visualizations provide researchers with transparent provenance information supporting reproducible research, while administrators gain impact analysis capabilities showing downstream dependencies when considering dataset modifications. Operational workshops train MassTech teams on scanner configuration, lineage interpretation for compliance reporting, and ongoing maintenance procedures ensuring lineage remains accurate as the platform evolves.

Advanced Financial Intelligence with Apptio

Phase 2 transitions the Commons Credits System from foundational ledger to sophisticated financial intelligence platform through Apptio's advanced analytics capabilities. IBM Expert Labs configures resource-based costing models mapping Snowflake compute costs—including warehouse usage, storage consumption, and data transfer—to individual user credit debits with granular precision based on QUERY_HISTORY data correlation. The implementation establishes tiered crediting structures rewarding high-quality dataset contributions validated via Data Quality scores with premium credit multipliers, while standard contributions receive base rates, creating economic incentives for data curation excellence.

Apptio chargeback models enable transparent cost pass-through for third-party AI services accessed via API Connect, automatically converting vendor costs to credit debits visible in user transaction histories. IBM configures multiple dashboard types serving distinct stakeholder needs: executive dashboards provide real-time visualization of credit circulation, system-wide usage patterns, financial sustainability metrics, and resource optimization opportunities for MassTech leadership; user portals embedded in the SI's web application via API Connect display personalized credit balances, detailed transaction histories, earning opportunities based on user capabilities, and spending projections helping researchers plan resource allocation; compliance dashboards generate automated audit reports linking every credit transaction to immutable PostgreSQL logs with complete provenance trails supporting transparency requirements and third-party verification.

The implementation includes optimization and forecasting capabilities leveraging historical usage patterns to predict future credit consumption, enabling proactive capacity planning and budget allocation. IBM configures efficiency incentive algorithms identifying opportunities like off-peak usage or resource consolidation, automatically awarding credit bonuses encouraging sustainable behaviors. Dynamic pricing models adjust credit-to-resource conversion rates based on platform utilization, demand elasticity, and strategic priorities, ensuring the Commons Credits System remains financially sustainable while promoting equitable access across diverse user communities.

Phase 2 Deliverables:

- Data Quality framework with 20+ rules monitoring dataset trustworthiness and automated remediation workflows
- Manta Data Lineage configuration tracking 3+ priority data sources with graphical visualizations accessible via API
- Enhanced Apptio dashboards providing executive insights, personalized user portals, and comprehensive compliance reporting
- Optimized credit earning and spending workflows with quality-based tiered rates and efficiency incentive algorithms
- Integration documentation enabling SI to embed quality indicators, lineage visualizations, and credit analytics into user interfaces

Phase 3: Optimization and Validation

IBM Expert Labs focuses Phase 3 on performance optimization, compliance validation, and comprehensive knowledge transfer ensuring MassTech can independently operate and evolve the Data Commons Collaborative as a sustainable innovation platform.

Strategic Platform Assessment

IBM conducts a one-day Data Platform Accelerator consulting engagement aligning Phase 1-2 implementations with evolving MassTech requirements and identifying optimization opportunities. The assessment reviews Data Catalog, Quality, and Lineage capabilities against original objectives, measuring actual performance against success criteria including metadata completeness, quality threshold compliance, and lineage accuracy. IBM evaluates Commons Credits System performance through multiple dimensions: transaction throughput analysis ensuring sub-second processing latency; audit trail integrity validation confirming 99.9%+ ledger accuracy; user satisfaction surveys measuring adoption barriers and feature requests; financial sustainability modeling projecting long-term platform economics.

The engagement validates advanced analytics use cases leveraging the established governance framework, prioritizing opportunities like AI model registry integration, synthetic data generation tracking, and cross-institutional collaboration metrics. IBM defines architecture enhancements supporting future expansion including additional data source connectors, federated query capabilities across distributed datasets, and advanced security features like confidential computing for sensitive data analysis. Migration strategy recommendations guide transitioning pilot deployments to production scale, addressing capacity planning, disaster recovery validation, and operational maturity requirements ensuring the platform can support Massachusetts' growing AI research community.

Performance Optimization and Refinement

IBM Expert Labs conducts comprehensive performance tuning based on real-world usage patterns observed during Phase 1-2 operations. Metadata quality enhancement iterates on Data Catalog enrichment rules incorporating user feedback from researchers, expanding business term glossaries with domain-specific vocabularies and refining classification schemes improving dataset discoverability. The team expands data protection rules addressing newly identified sensitivity requirements and adjusts quality thresholds based on observed error rates and remediation effectiveness.

Lineage accuracy validation conducts additional Lineage Reviews as new data sources onboard, verifying compliance with EU AI Act traceability requirements and Massachusetts data protection regulations. IBM configures supplementary scanners capturing previously unmapped data flows and validates lineage visualizations against actual data movement patterns through Snowflake audit logs. API performance tuning optimizes IBM API Connect configurations reducing latency for high-frequency operations like catalog searches and credit balance queries, implementing caching strategies for frequently accessed datasets and connection pooling for database interactions.

Credit system refinement adjusts Apptio allocation models based on observed usage patterns, recalibrating earning rates ensuring adequate credit supply for active contributors and spending rates maintaining financial sustainability while preventing resource hoarding. IBM analyzes efficiency incentive effectiveness, modifying bonus structures amplifying desired behaviors like off-peak usage or data quality improvement. The team validates rate limiting policies through load testing, ensuring API Connect can handle projected growth in user adoption and dataset volume without degradation.

FAIR Principles Validation and Compliance

IBM conducts formal validation against FAIR principles ensuring the Data Commons meets international standards for data sharing excellence. Findability assessment verifies that datasets are assignable unique persistent identifiers, metadata is indexed in searchable resources accessible via IBM Knowledge Catalog APIs, and metadata includes clear data identifiers linking to source systems. Accessibility validation confirms that datasets and metadata are retrievable through standardized protocols (HTTPS, APIs), authentication and authorization procedures are clearly documented through IBM API Connect developer portals, and metadata remains accessible even when datasets are restricted.

Interoperability assessment validates that metadata uses formal, accessible, shared vocabularies aligned with Dublin Core, Schema.org, and DCAT standards, metadata includes qualified references to other datasets visible through Manta Data Lineage, and APIs follow RESTful design principles enabling integration with diverse research tools. Reusability validation ensures metadata provides rich descriptions including data quality scores, provenance information, and usage licenses, datasets meet community standards documented in governance policies, and clear usage guidelines specify permitted applications through Commons Credits System terms of service.

IBM documents compliance evidence supporting audit requirements, generating formal attestation reports MassTech can present to funding agencies, regulatory bodies, and research partners demonstrating adherence to international data sharing standards.

Knowledge Transfer and Sustainability

Comprehensive enablement ensures MassTech can independently operate and evolve the platform beyond IBM's implementation engagement. IBM delivers multi-day administrator training workshops covering Data Catalog curation procedures including metadata enrichment, business glossary management, and access control configuration; Data Quality rule development, SLA workflow customization, and remediation process optimization; Lineage scanner deployment for new data sources, interpretation of complex lineage visualizations, and compliance reporting automation; API Connect policy updates, gateway configuration, and developer portal customization; Apptio dashboard development, cost model adjustments, and credit rate optimization.

Developer documentation provides complete API reference guides with code examples in multiple languages, integration patterns for common use cases like dataset submission and credit balance checking, troubleshooting procedures addressing frequent issues, and performance optimization techniques. Operational runbooks detail routine maintenance procedures including database backup verification, log rotation, and certificate renewal; incident response protocols with severity classification, escalation paths, and communication templates; capacity management guidelines projecting storage growth, compute scaling triggers, and performance monitoring; disaster recovery procedures validating backup integrity, recovery time objectives, and business continuity planning.

Change management support facilitates ongoing user adoption through stakeholder communication campaigns highlighting platform capabilities and success stories, feedback collection mechanisms including user surveys, focus groups, and analytics on feature usage, and continuous improvement processes incorporating community input into platform evolution roadmaps. IBM establishes support transition procedures connecting MassTech administrators with IBM Expert Labs resources during an

initial stabilization period, gradually transferring operational responsibility while maintaining advisory availability for complex technical issues.

Phase 3 Deliverables:

- Platform optimization report documenting performance improvements, architectural refinements, and capacity recommendations
- FAIR principles compliance validation with formal attestation supporting audit and funding requirements
- Comprehensive training materials including video tutorials, hands-on lab exercises, and certification assessments
- Operational runbooks covering maintenance procedures, incident response protocols, and capacity management guidelines
- Long-term sustainability plan with recommendations for scaling, community growth strategies, and continuous improvement frameworks
- Support transition documentation enabling seamless handoff from implementation to operations

This implementation approach leverages IBM's proven methodologies, Apptio's financial intelligence platform, and deep Massachusetts expertise to deliver a future-ready Data Commons Collaborative that positions the Commonwealth as a national leader in responsible AI innovation, fostering cross-sector collaboration while maintaining the highest standards of governance, security, and ethical oversight.

2025 IBM CORPORATION | CLIENT AND IBM CONFIDENTIAL

4. Integration with Chosen SI's PMO

Integrated Solution Delivery Through Strategic Partnership

IBM's collaboration with the selected System Integrator (SI) and Snowflake creates a cohesive enterprise architecture where each partner contributes specialized expertise toward a unified Data Commons Collaborative. IBM provides the governance intelligence layer through its watsonx platform suite—including watsonx.data Intelligence for metadata management, watsonx.governance for AI ethics and compliance, and watsonx.ai for synthetic data generation—while Snowflake delivers the scalable cloud-native data processing foundation with its elastic compute capabilities and secure storage infrastructure. The SI orchestrates these components through system integration and user-facing portal development, creating seamless experiences for researchers, government agencies, and private sector innovators. IBM API Connect serves as the critical integration gateway, securing all interactions between the SI's web portal and backend systems with OAuth 2.0 authentication, rate limiting, and comprehensive audit logging. This architecture enables the SI to focus on user experience design and workflow optimization while IBM ensures backend reliability, data provenance through Manta Data Lineage, and ethical AI practices through continuous bias monitoring across 120+ demographic variables.

Phased Implementation and PMO Coordination

The three-phase delivery model demonstrates how IBM Expert Labs integrates with the SI's Project Management Office to ensure coordinated milestone achievement and risk mitigation. In Phase 1, IBM deploys the foundational IBM PostgreSQL ledger with immutable audit triggers by Week 6 and integrates API Connect for secure endpoints by Week 9, feeding usage data from Snowflake's QUERY_HISTORY and ACCESS_HISTORY views into the governance framework. Simultaneously, the SI develops the initial web portal framework and user authentication system, consuming IBM's catalog APIs to enable dataset discovery through IBM Knowledge Catalog. Phase 2 introduces IBM Apptio for sophisticated credit management analytics by Week 16 and full watsonx.governance integration for AI fairness evaluation by Week 20, with the SI enhancing portal functionality to display real-time credit balances, fairness metrics, and dataset lineage visualizations. Throughout Phase 3, IBM refines metadata quality and lineage accuracy based on SI feedback while the SI rolls out public dataset access and administrative tools. IBM Expert Labs practitioners embed within the SI's PMO structure—contributing governance analysts to maintain RAID logs with watsonx.governance risk assessments, supporting Scrum Masters for Agile workstreams using watsonx.ai, and managing Waterfall deliverables like compliance validation—ensuring that Snowflake's scalable infrastructure supports both rapid iteration and structured compliance milestones.

Technical Integration and API-Driven Collaboration

IBM's API-first integration strategy enables the SI to build flexible, user-centric interfaces while maintaining enterprise-grade security and governance. The technical integration specifications define five core API endpoints secured by IBM API Connect: /api/v1/catalog/search for full-text and semantic search across IBM Knowledge Catalog, /api/v1/assets/{id}/metadata for detailed dataset metadata retrieval, /api/v1/lineage/{asset_id} for graphical data provenance via Manta, /api/v1/credits/balance/{user_id} for real-time credit transactions managed by Apptio, and /api/v1/governance/bias/scan to initiate fairness evaluations through watsonx.governance. These APIs follow RESTful design principles with consistent response structures, enabling the SI to create reusable UI components—such as dataset cards displaying descriptive metadata from IKC, lineage diagrams showing data flows from external sources through Snowflake transformations, and fairness dashboards presenting disparate impact metrics and statistical

parity differences. IBM provides OpenAPI specifications for automatic client generation, comprehensive integration guides, and a full-featured sandbox environment, while IBM Expert Labs engineers maintain dedicated support channels for troubleshooting. Snowflake's role as the unified data repository ensures that whether datasets originate from Hugging Face, World Bank, or Massachusetts state agencies, they flow through consistent ingestion pipelines—profiled by IBM's automated metadata extraction, stored in Snowflake's versioned tables with time-travel capabilities, and cataloged in IKC with appropriate business terms and classifications—before becoming discoverable through the SI's search interfaces.

Governance, Compliance, and Continuous Improvement

The partnership ensures robust governance through complementary capabilities: IBM watsonx.governance enforces FAIR principles and regulatory compliance (HIPAA, EU AI Act, Massachusetts data protection regulations) through customizable workflows with role-based access controls, while Snowflake provides tamper-evident audit logs and encryption at rest and in transit, and the SI implements user authentication with multi-factor authentication and session management. IBM Guardium Data Discovery scans unstructured assets for PII/PHI, feeding classification findings into watsonx.governance risk scoring questionnaires that route high-sensitivity datasets through approval workflows before the SI's portal grants access. For the Commons Credits System, IBM PostgreSQL maintains the immutable ledger with database triggers recording every transaction, while Snowflake supplies raw usage metrics that IBM watsonx Orchestrate processes into credit debits and credits, and IBM Apptio generates analytics dashboards that the SI embeds into administrative interfaces. IBM Expert Labs supports the SI's change management initiatives by developing training modules on AI fairness concepts, governance workflows, and credit system mechanics—delivered through watsonx.governance's educational features and hosted in the SI's Confluence knowledge repository. Regular joint technical synchronization meetings, shared success metrics around platform adoption and compliance validation, and iterative usability testing ensure continuous refinement based on real-world usage patterns from Massachusetts' diverse AI research community, with IBM providing governance intelligence, Snowflake enabling scalable data operations, and the SI delivering intuitive experiences that make sophisticated capabilities accessible to users regardless of technical expertise.

2025 IBM CORPORATION | CLIENT AND IBM CONFIDENTIAL

5. IBM Public Sector References

IBM's proven track record in delivering data governance, AI innovation, and metadata management solutions for public sector clients demonstrates our capability to support the Massachusetts AI Hub Data Commons Collaborative. The following representative engagements showcase our expertise in implementing enterprise-grade platforms that balance technical sophistication with user accessibility while maintaining rigorous governance standards.

Wilson Center - Federal Government Data Quality and Governance

Organization: Woodrow Wilson International Center for Scholars, Washington, D.C.

Challenge: The Wilson Center's Earth Challenge project required robust data quality management for crowdsourced environmental datasets collected through mobile applications across diverse global communities. The organization needed automated metadata cataloging with appropriate access controls and governance frameworks, eliminating manual image labeling requirements for thousands of submissions. Additionally, the Center sought solutions for interoperability challenges in metadata harmonization across heterogeneous data sources from multiple countries and scientific disciplines.

IBM Solution: IBM deployed watsonx.data Intelligence with IBM Knowledge Catalog to establish automated data quality assessment capabilities. The implementation featured an unsupervised and semi-supervised image recognition model detecting reliable imagery among thousands of submissions—a task impossible to complete manually at scale. IBM configured searchable catalogs enabling researchers to access image data with comprehensive AI factsheets and data nutrition labels displayed in intuitive dashboards. Leveraging IBM Cloud Pak for Data, the solution incorporated a first-of-its-kind Data Reliability Index combining authenticity, usability, label confidence, and image relevance on a 0-100 scale, automatically removing invalid data while tracking quality improvements over time.

Outcomes:

- Developed predictive models forecasting litter volumes five years into the future, supporting policy planning and resource allocation decisions
- Created end-to-end AI lifecycle management with custom digital dashboards making environmental data accessible to diverse stakeholder communities including researchers, policymakers, and advocacy organizations
- Empowered United Nations Environment Programme stakeholders, including Costa Rica's environmental agencies, to track progress toward national goals of eliminating single-use plastics through data-driven evidence
- Established governance frameworks ensuring crowdsourced data meets scientific standards for reliability, reproducibility, and ethical use in environmental policy development

Relevance to Data Commons Collaborative: The Wilson Center engagement demonstrates IBM's capability to implement automated metadata management, data quality frameworks, and governance controls for large-scale collaborative data platforms—directly applicable to MassTech's requirements for curating diverse datasets, ensuring FAIR principles compliance, and enabling transparent data quality assessment supporting responsible AI research across academic, government, and private sector communities.

Sonoma County, California - Integrated Data Governance for Social Services

Organization: Sonoma County Central Information Technology, California

Challenge: Sonoma County operated fragmented data systems across multiple social service departments, preventing holistic understanding of client needs and coordinated support delivery. The County required seamless data integration enabling caseworkers to access comprehensive client information from siloed systems while maintaining privacy protections and regulatory compliance. User-friendly interfaces accessible to non-technical frontline workers were essential for adoption and effective service delivery targeting homelessness reduction.

IBM Solution: IBM implemented data integration capabilities transforming business processes to enable secure data sharing between departments and coordinated support between service providers and clients. The solution consolidated information from disparate source systems, providing 360-degree holistic views of client interactions, service histories, and support needs accessible through intuitive dashboards designed for caseworkers without technical expertise.

Outcomes:

- Achieved 9% reduction in homelessness through data-driven coordination of housing assistance, mental health services, employment support, and substance abuse treatment programs
- Enabled real-time access to comprehensive client information supporting informed decision-making by frontline caseworkers during critical intervention moments
- Delivered user-friendly interfaces empowering non-technical staff to quickly locate information through dashboards and individual client records without specialized training requirements
- Established reliable, stable platform supporting continuous operations without system disruptions, building trust among caseworkers and administrators

Client Testimonial: *"To collaborate, you need a way of getting data out of all these systems in a format that is very user-friendly for the caseworkers. That's where IBM came in,"* says Carolyn Staats, Director of Innovation at Sonoma County Central IT. *"It's very user friendly for non-technical people. You can quickly get the information you're looking for, whether a dashboard or individual client information. The solution works and it doesn't break."*

Relevance to Data Commons Collaborative: Sonoma County's successful deployment illustrates IBM's expertise in creating accessible, user-friendly data platforms serving diverse user communities with varying technical capabilities—mirroring MassTech's requirement for intuitive interfaces enabling researchers, government agencies, and private sector innovators to discover and utilize datasets regardless of technical expertise. The emphasis on reliability, stability, and governance for sensitive social services data directly addresses Data Commons requirements for secure handling of potentially sensitive research datasets while maintaining operational excellence.

VIA Metropolitan Transit, San Antonio, Texas - AI Implementation with Expert Labs Support

Organization: VIA Metropolitan Transit Authority, San Antonio, Texas

Challenge: VIA Metropolitan Transit sought to enhance digital customer experiences through AI-powered virtual assistance while building internal team capabilities for sustained innovation. The organization

required an automated digital agent drawing on call center data to answer frequently asked questions, interact with third-party APIs for real-time transit information and directions, and converse in both English and Spanish supporting San Antonio's diverse community. Operational analytics, cloud-based architecture, and comprehensive knowledge transfer to VIA's development team—undertaking their first AI project—were critical success factors.

IBM Expert Labs Solution: IBM Expert Labs collaborated closely with VIA's development team through structured weekly planning sessions, hands-on technical mentorship, and knowledge sharing ensuring sustainable in-house capabilities. IBM practitioners provided guidance on AI model development, API integration patterns, multilingual natural language processing, and operational analytics implementation while encouraging VIA developers to take ownership of components aligned with their growing expertise. The collaborative approach emphasized skill development alongside technology delivery, with IBM experts maintaining availability for troubleshooting and advanced technical challenges.

Outcomes:

- Successfully deployed bilingual virtual assistant handling common inquiries, reducing call center volume while improving response times and customer satisfaction metrics
- Built internal AI development capabilities enabling VIA's team to independently sustain, enhance, and expand the virtual assistant as ridership needs evolve
- Established collaborative partnership model where IBM Expert Labs served as trusted advisors rather than external implementers, fostering knowledge transfer and team confidence
- Created foundation for future AI initiatives across VIA's operations through upskilled development team and proven implementation methodologies

Client Testimonial: *"AI and digital agents were new to our team and to our developers. We wanted to create something that our whole team could support, sustain and grow, and have a partner that would do it along with us. That's exactly what we gained with IBM,"* says Steve Young, Vice President of Technology and Innovation. *"There were weekly planning calls about things the IBM team was working on and things our team handled. We had assigned some junior developers and we wanted them to grow. The IBM experts were really great about sharing, saying, 'Hey, if you get stuck or need help, please reach out to us.'"*

Relevance to Data Commons Collaborative: VIA Metropolitan Transit's experience demonstrates IBM Expert Labs' commitment to collaborative implementation, comprehensive knowledge transfer, and building sustainable internal capabilities—directly aligned with MassTech's requirement for administrator training, operational documentation, and support transition enabling independent platform operation. The emphasis on mentorship, accessibility to technical expertise, and empowering client teams to evolve solutions reflects IBM's approach to the Data Commons implementation, ensuring MassTech administrators can confidently manage, enhance, and scale the platform beyond initial deployment while maintaining access to IBM Expert Labs advisors for complex challenges.

Reference Alignment with Data Commons Requirements

These representative engagements demonstrate IBM's proven capabilities across critical Data Commons requirements:

Data Governance and Quality: Wilson Center's automated data quality assessment, reliability indexing, and metadata harmonization directly parallel Data Commons needs for ensuring dataset trustworthiness, FAIR principles compliance, and quality-driven credit allocation.

User-Centric Design: Sonoma County's emphasis on user-friendly interfaces accessible to non-technical users reflects MassTech's requirement for intuitive experiences serving researchers, government agencies, and private sector innovators with diverse technical capabilities.

Knowledge Transfer and Sustainability: VIA Metropolitan Transit's collaborative implementation with embedded knowledge transfer illustrates IBM Expert Labs' commitment to building client capabilities, comprehensive training, and support transition—ensuring MassTech can independently operate and evolve the Data Commons platform.

Public Sector Expertise: All three engagements showcase IBM's understanding of public sector governance requirements, regulatory compliance constraints, stakeholder engagement complexities, and mission-driven outcomes prioritizing societal benefit over commercial metrics.

Scalable, Reliable Infrastructure: Sonoma County's stable, continuously operating platform and Wilson Center's global-scale crowdsourced data management demonstrate IBM's capability to deliver enterprise-grade reliability supporting critical public services and research initiatives.

These proven successes position IBM to deliver a Data Commons Collaborative that meets MassTech's vision for responsible AI innovation, transparent governance, and equitable access while building sustainable internal capabilities ensuring long-term platform success across the Commonwealth.

Customers require us to schedule a Webconference via their IBM Account Executive.

Reference	Name	Email	Main Phone Number
Wilson Center - Federal Government Data Quality and Governance Organization: Woodrow Wilson International Center for Scholars, Washington, D.C. (No specific individual named in the engagement; using project/organization lead contacts)	Ryan McKenna (Earth Challenge 2020 Project Contact) or General Inquiry	ryan.mckenna@wilsoncenter.org or fellowships@wilsoncenter.org	(202) 691-4217 or (202) 691-4000
Sonoma County, California - Integrated Data Governance for Social Services Organization: Sonoma County Central	Dan Fruchey	Not publicly available (Use general IT directory at permitsonoma.org or	(707) 565-1900

Information Technology, California Dan Fruchey Information Systems Director at County of Sonoma https://www.linkedin.com/in/dan-fruchey-220a955/ **Linkedin**		sonomacounty.ca.gov for inquiries)	
VIA Metropolitan Transit, San Antonio, Texas - AI Implementation with Expert Labs Support Organization: VIA Metropolitan Transit Authority, San Antonio, Texas	Steve Young (Vice President of Technology and Innovation)	syoung@viainfo.net	(210) 362-2000 (VIA main line; direct line not publicly available)

Customers require us to schedule a Webconference via their IBM Account Executive.

6. Conclusion

IBM is honored to propose **a solution uniquely positioned to address critical areas of the Data Commons Collaborative mission to centralize access to data for AI development, research, and innovation**. Through our strategic partnership with the selected System Integrator and data platform provider, we deliver specialized IBM technology capabilities that position the Commonwealth as a national leader in responsible, scalable AI innovation. Our governance intelligence, metadata management, API orchestration, and financial transparency solutions balance technical sophistication with accessibility, ensuring researchers, government agencies, and private sector innovators can collaborate effectively while maintaining the highest standards of security and ethical oversight.

IBM's contribution centers on four essential pillars complementing the System Integrator's platform development and data infrastructure capabilities. Our **watsonx platform suite**—including watsonx.data Intelligence for metadata management, watsonx.governance for AI fairness and compliance, and watsonx.ai for synthetic data generation—integrates seamlessly with the data platform's scalable cloud-native infrastructure and the SI's user-centric portal development. **IBM API Connect** serves as the secure orchestration gateway, enabling controlled, auditable interactions across all platform components while supporting the **Commons Credits System** through IBM PostgreSQL's immutable ledger and IBM Apptio's sophisticated analytics capabilities.

IBM's deep commitment to Massachusetts extends beyond technology delivery. With over 2,000 professionals across our Lowell, Cambridge, and Boston-area facilities, including the IBM-MIT Innovation Center in Kendall Square, we bring unparalleled local expertise and rapid response capabilities. Our extensive public sector experience—spanning Massachusetts state agencies, federal entities, and global enterprises—demonstrates our ability to navigate complex regulatory environments while delivering transformative outcomes. The IBM Expert Labs team will embed within the SI's Project Management

Office, contributing governance analysts, technical specialists, and change management expertise to ensure seamless coordination across all implementation phases.

Our phased delivery approach minimizes risk while maximizing value. Phase 1 establishes the foundational governance and credit infrastructure, Phase 2 introduces advanced analytics and AI fairness capabilities, and Phase 3 optimizes performance and validates FAIR principles compliance. Throughout this journey, IBM will work collaboratively with MassTech leadership, the SI, and Snowflake to adapt to evolving requirements, incorporate stakeholder feedback, and ensure the platform grows organically with the Commonwealth's AI innovation ecosystem.

The technical architecture we propose—built on API-first integration principles, separation of concerns, and progressive delivery—ensures long-term sustainability and extensibility. Whether researchers require curated datasets for machine learning, government agencies need secure data sharing with compliance frameworks, or private innovators seek synthetic data for privacy-preserving analytics, our solution provides enterprise-grade capabilities accessible through intuitive interfaces. IBM's leadership in AI governance, evidenced by our recognition in IDC's MarketScape and Gartner's evaluations, guarantees that Massachusetts benefits from cutting-edge technology aligned with emerging regulations including the EU AI Act and NIST AI Risk Management Framework.

We are excited about the opportunity to partner with the Massachusetts Technology Collaborative in realizing the vision established by the AI Strategic Task Force and the Healey-Driscoll administration. IBM's combination of advanced technology platforms, deep implementation expertise through Expert Labs, strategic alliances with industry leaders like Snowflake, and unwavering commitment to the Commonwealth positions us uniquely to deliver a Data Commons Collaborative that accelerates AI innovation while upholding the values of transparency, fairness, and public benefit. Together, we will build a platform that not only meets today's requirements but adapts seamlessly to tomorrow's opportunities, cementing Massachusetts' position at the forefront of responsible AI leadership.

2025 IBM CORPORATION | CLIENT AND IBM CONFIDENTIAL

Appendix – Pricing

IBM Technology and Services Investment Summary

Massachusetts AI Hub Data Commons Collaborative

RFP No. 2026-JAII-01

Investment Overview

IBM proposes a comprehensive three-year solution for the Massachusetts AI Hub Data Commons Collaborative, encompassing IBM Technology software licenses and IBM Expert Labs implementation services. This investment establishes enterprise-grade governance, metadata management, AI fairness capabilities, and the Commons Credits System supporting MassTech's vision for responsible AI innovation.

Year 1 Investment Summary

IBM Technology Software - Year 1

Component	SKU	Quantity	Investment
Watsonx.data Intelligence	D10LXZX	2,500	$165,600.00
Watsonx.data	D0HLLZX	9	$8,280.00
Cloud Pak for Data - PostgreSQL	D01CFZX	200	$2,400.00
Watsonx.governance	D0FHQZX	6,750	$74,520.00
Watsonx.ai (Synthetic Data Generation)	D0FHQZX	167	$1,840.00
Watsonx.orchestrate	D0B3YZX	167	$5,520.00
Apptio Standard	D0ICKZX	15	$118,800.00
Apptio All Education Pass	D0GWJZX	2	$7,704.00
IBM Cloud Pak for Integration - API Calls Resource Value Unit	D052XZX	30	$16,632.00
Year 1 Software Subtotal			**$401,296.00**

IBM Expert Labs Services - Year 1

Service Description	SKU	Quantity	Investment
Expert Labs - Data Catalog Build	D12DIZX	1	$40,000.00
Expert Labs - Data Quality Build	D10DTZX	1	$50,000.00
Expert Labs - Lineage Build	D10DTZX	1	$50,000.00

2025 IBM CORPORATION | CLIENT AND IBM CONFIDENTIAL

IBM Expert Labs AI Governance One Week Remote Solution Build - SaaS Engagement	D070BZX	4	$40,000.00
IBM Expert Labs Build AI Governance - SaaS Engagement Remotely Delivered SW Services	D0GC7ZX	5	$86,500.00
IBM Expert Labs Advise AI Governance - SaaS Engagement Remotely Delivered SW Services	D0GC9ZX	2	$38,200.00
IBM Expert Labs Advise AI Assistants SaaS Engagement Remotely Delivered Service	D06ZVZX	2	$38,200.00
IBM Expert Labs Build AI Assistants SaaS Engagement Remotely Delivered	D06ZWZX	3	$51,900.00
Expert Labs Implementation	D0IG9ZX	2	$158,000.00
Expert Labs Managed Services	D0HGAZX	1	$52,800.00
IBM Expert Labs Install Integration - Non-SaaS Engagement Remotely Delivered SW Services	D02DAZX	3	$51,900.00
IBM Exp Labs - IBM Cloud Pak for Integration One Week Soln Arch Non-SaaS Engagement Remotely Delivered SW Services	D02DBZX	2	$45,000.00
IBM Expert Labs Build Integration - Non-SaaS Engagement Remotely Delivered SW Services	D02DCZX	4	$69,200.00
Year 1 Expert Labs Subtotal			**$771,700.00**

Year 1 Total Investment: $1,172,996.00

Year 2 Investment Summary

IBM Technology Software - Year 2

Component	SKU	Quantity	Investment
Watsonx.data Intelligence	D10LXZX	2,500	$182,160.00
Watsonx.data	D0HLLZX	9	$9,108.00
Cloud Pak for Data - PostgreSQL	D01CFZX	200	$2,640.00
Watsonx.governance	D0FHQZX	6,750	$81,972.00
Watsonx.ai (Synthetic Data Generation)	D0FHQZX	167	$2,024.00
Watsonx.orchestrate	D0B3YZX	167	$6,072.00
Apptio Standard	D0ICKZX	15	$130,680.00
Apptio All Education Pass	D0GWJZX	2	$8,474.40
IBM Cloud Pak for Integration - API Calls Resource Value Unit	D052XZX	150	$83,160.00
Year 2 Software Subtotal			**$506,290.40**

Year 2 Total Investment: $506,290.40

Note: Year 2 includes software license renewals with standard annual escalation. No additional Expert Labs services are anticipated in Year 2 as implementation is completed in Year 1.

Year 3 Investment Summary

IBM Technology Software - Year 3

Component	SKU	Quantity	Investment
Watsonx.data Intelligence	D10LXZX	2,500	$200,376.00
Watsonx.data	D0HLLZX	9	$10,018.80
Cloud Pak for Data - PostgreSQL	D01CFZX	200	$2,904.00
Watsonx.governance	D0FHQZX	6,750	$90,169.20
Watsonx.ai (Synthetic Data Generation)	D0FHQZX	167	$2,226.40
Watsonx.orchestrate	D0B3YZX	167	$6,679.20
Apptio Standard	D0ICKZX	15	$143,748.00
Apptio All Education Pass	D0GWJZX	2	$9,321.84
IBM Cloud Pak for Integration - API Calls Resource Value Unit	D052XZX	150	$91,476.00
Year 3 Software Subtotal			**$556,919.44**

Year 3 Total Investment: $556,919.44

Note: Year 3 includes software license renewals with standard annual escalation. No additional Expert Labs services are anticipated in Year 3.

Three-Year Total Investment Summary

Year	Software Investment	Services Investment	Annual Total
Year 1	$401,296.00	$771,700.00	**$1,172,996.00**
Year 2	$506,290.40	$0.00	**$506,290.40**
Year 3	$556,919.44	$0.00	**$556,919.44**
Three-Year Total	**$1,464,505.84**	**$771,700.00**	**$2,236,205.84**

Investment Breakdown by Category

IBM Technology Software (3-Year Total: $1,464,505.84)

- **Governance & Metadata Management:** Watsonx.data Intelligence, Watsonx.data, IBM Knowledge Catalog
- **AI Fairness & Ethics:** Watsonx.governance for bias detection, compliance validation, and continuous monitoring

- **Synthetic Data Generation:** Watsonx.ai supporting privacy-preserving analytics and AI model training
- **Commons Credits System:** IBM PostgreSQL for immutable ledger, Watsonx.orchestrate for transaction workflows, Apptio for financial intelligence
- **API Orchestration:** IBM Cloud Pak for Integration (API Connect) for secure gateway and developer portal

IBM Expert Labs Services (Year 1 Total: $771,700.00)

- **Data Governance Implementation:** Data Catalog Build, Data Quality Build, Lineage Build
- **AI Governance & Fairness:** AI Governance solution builds, advisory services, and AI assistants implementation
- **Integration & API Management:** Integration installation, architecture services, and API Connect deployment
- **Implementation & Managed Services:** Full platform implementation and initial managed services support

Payment Terms and Conditions

Payment Schedule:

- Year 1: Payable upon contract execution and delivery milestones per Phase 1-3 implementation schedule
- Year 2: Annual software renewal payable at contract anniversary
- Year 3: Annual software renewal payable at contract anniversary

IBM is a vendor on the Commonwealth of Massachusetts' ITS75 Software and Services contract, a multi-state agreement that allows government entities to purchase software, software-as-a-service (SaaS), platform-as-a-service (PaaS), and infrastructure-as-a-service (IaaS). This contract may be leveraged by customer. Also, the customer can leverage the IBM NASPO Cloud Contract. Additional options are available here - https://www.carahsoft.com/ibm/contracts.

Scope Modifications:

- Any scope changes requiring additional software licenses or services will be documented through change order process
- Additional Expert Labs services available at standard rates: $1,800-2,200 per day depending on specialty and seniority

Investment Value Proposition

This comprehensive investment delivers:

✓ **Enterprise-Grade Platform:** Proven IBM watsonx and industry-leading data technology for a secure and collaborative environment

✓ **Comprehensive Implementation:** IBM Expert Labs services ensuring successful deployment, knowledge transfer, and operational readiness

✓ **Scalable Architecture:** Cloud-native infrastructure supporting Massachusetts' growing AI research community

✓ **Governance Excellence:** Industry-leading AI fairness, metadata management, and compliance capabilities

✓ **Financial Transparency:** Commons Credits System with immutable audit trails and sophisticated analytics

✓ **Long-Term Sustainability:** Training, documentation, and support transition enabling independent platform operation

✓ **Proven ROI:** Reference implementations demonstrating measurable outcomes in data quality, governance efficiency, and user adoption

For questions regarding pricing, payment terms, or scope, please contact:

Shawn Noyes
IBM Technology Sales Lead
Commonwealth of Massachusetts and Higher Education
Phone: 617.901.5848
Email: shawn.noyes@ibm.com

Pricing valid through December 31, 2025. All prices in USD. Standard IBM terms and conditions apply.

Appendix – Description of Services

Consolidated Service Descriptions for IBM Expert Labs Services

This document compiles all the service descriptions extracted from the provided documents. It is organized by individual service or scope document for clarity. Each section includes the full text of the service description, scope, activities, assumptions, and other relevant details. This consolidated version

can be used as a basis for incorporation into an RFP. Note that original formatting (e.g., tables, images) has been adapted to plain text for this compilation.

1. IBM Expert Labs Architect MQ (i.e., IBM API Connect) - Scope Document (from "Architect MQ - Scope Document.docx")

Guidance for Use of this Document

Note: Use this Guidance page to determine if this SCOPE DOCUMENT fits your needs. Discard it prior to presenting the document.

Purpose: Set expectations with the Customer & Delivery on how a Services Part order will be used.

Description: The Scope Document is for use for IBM Cloud and Cognitive Expert Labs Services engagements where the Client wants a custom solution delivered through the sale of part numbers / PIDs. In addition, this form may be used against a Tile.

Note: The Scope document is not legally binding and does not require client acknowledgment response via email or signature.

IBM's Architect API Connect service.

The goal of this Architect API Connect engagement is to assist the Client with the following:

- Creation of a Blueprint document for the Client's API Connect Solution.

Description	Quantity	Link to Service Description (Optional)
IBM Expert Labs Architect API Connect - Non-SaaS OR IBM Expert Labs Architect API Connect - SaaS	10	IBM API Connect Service Description

Estimated Schedule

These Services will be performed consistent with the estimated schedule mutually agreed upon by both IBM and the Client. Both the Client and IBM agree to make reasonable efforts to carry out our respective responsibilities according to the estimated schedule.

Estimated Start Date: [Month, day, year]

Estimated End Date: [Month, day, year]

Work Activities

IBM will perform the following activities over the course of the entire project:

Activity 1 - Solution Management

IBM will provide solution management for the IBM responsibilities in this Scope Document. The purpose of this activity is to provide technical direction and control of IBM project personnel and to provide a framework for project planning, communications, reporting, procedural, and contractual activity. This activity is composed of the following tasks:

Planning IBM will:

- Review the Service Description and the contractual responsibilities of both parties with the Client's Solution Manager;
- Maintain project communications through the Client's Solution Manager;
- Coordinate the establishment of the solution environment;
- Assist the Client's Solution Manager to prepare and maintain the project plan for the performance of this Service Description, which will include the activities, tasks, assignments, and estimates; and
- Review with the Client any hardware required for the performance of this Service Description.

Project Tracking and Reporting IBM will:

- Review project tasks, schedules, and resources and make changes or additions, as appropriate. Measure and evaluate progress against the project plan with the Client's Solution Manager;
- Review the IBM standard invoice format and billing procedure to be used on the project, with the Client's Solution Manager;
- Work with the Client's Solution Manager to address and resolve deviations from the project plan;
- Conduct regularly scheduled project status meetings;
- Report time spent on the project to the Client's Solution Manager; and
- Coordinate and manage the technical activities of IBM project personnel.

Activity 2 – Discovery Sessions and Blueprint Creation

Under this activity, IBM will provide an API Connect architect for 2 weeks to lead a series of discovery sessions with the Client's business sponsor, subject matter experts (SMEs), and application owners. This may include whiteboard sessions, interviews with Client staff, assessment of existing systems, and review of documentation to create and present a Blueprint for building and deploying API Connect based on IBM recommended practices. Tasks will typically include:

a. Establish a clear understanding of the Client's business priorities and goals;

b. Understand the Client's existing environment;

c. Identify challenges and pain points;

d. Define the Client's timelines and key milestones;

e. Detailed technical sessions reviewing:

f. Infrastructure & form factors;

g. High availability & disaster recovery;

h. Security & networking;

i. API Connect configuration & topology (including Provider Organizations, Consumer Organizations, Catalogs, and Spaces);

j. Release management (DevOps), CI/CD pipelines, monitoring, logging, and Developer Portals;

k. Identify initial perceived risks and possible mitigations;

l. Define key pre-requisites for solution;

m. Define scope and plan for building solution; Prepare final Blueprint report;

o. Present Blueprint to Client.

Assumptions

- Normal work hours will be 8:30am – 5:00pm, Monday through Friday, in the local time of the location where the work will be performed.
- If travel and living expenses are deemed necessary to complete this Scope Document, IBM will charge the Client actual costs. Travel and living costs are not included in this Scope Document and are in addition to the services charges. Travel and living costs, if deemed necessary by both parties, will be documented and mutually agreed to via a separate Statement of Work (SOW).
- Any changes to this Scope Document will require mutual agreement by IBM and the Client.
- The processes used by the resources are provided and monitored by the Client, including controls regarding access, security, encryption, use and transmission of data, backup and recovery of the database, and any stored data.
- IBM does not commit to any service levels.
- The following items are not included as part of this engagement: a. Hardware installation; b. Creation of virtual machines and operating systems; c. Backup, recovery, and disaster recovery; d. High availability setup and testing; e. Security compliance and hardening; f. Governance processes; g. Defining the functional testing criteria; h. Performance testing; i. Deployment automation; j. Test automation.

2. IBM Technology Expert Labs Build Data Catalog Build Data Catalog (SaaS).pdf Service Description

IBM Technology Expert Labs Build Data Catalog (SaaS)

This Service Description describes the Acceleration Services to support Client's use of an IBM Cloud Service. The applicable order documents provide pricing and additional details about Client's order. Upon acceptance of Client's order, this Service Description and the Cloud Services Agreement found at http://www.ibm.com/terms/csa (or equivalent agreement for Cloud Services between the parties) applies to Client's use of the Acceleration Services.

1. **Acceleration Services**

Acceleration Services are expert services, as described in this Service Description (SD) and are delivered remotely. For the purpose of this SD, if the base agreement references Cloud Services, then all such references apply to Acceleration Services, or Services, as applicable, and that base agreement shall govern the terms of this SD.

1.1 **Services**

The Client may select from the following available services.

1.1.1 IBM Technology Expert Labs Build Data Catalog (SaaS)

This Service provides Client with a fundamental configuration of the Data Catalog component of IBM watsonx.data intelligence. The fixed scope Service enables Client to establish a data governance driven data management foundation, using the IBM Knowledge Catalog, to assist with thorough metadata management, data discovery and governed data use.

The scope includes the following:

- Deploy a single cross-industry Knowledge Accelerator (KA).
- Define, create and customize one (1) governance workflow.
- Create and configure Data Asset Catalogs to support curated consumer access, global search, and relationship explorer to governed data.
- Perform metadata import (MDI) and metadata enrichment (MDE).

This offering includes a maximum of 120 hours of consultation and delivery with 12 hours of program management per Engagement and consists of the following activities:

- Create and configure Cloud Pak for Data users, groups & roles specific to the Data Catalog offering.
- Create and configure Platform Connections for two (2) existing Client data sources, with up to ten (10) tables total.
- Deploy with guidance - one (1) (KA), aligned with the Client specific industry and/or requirements.
- Define, create and customize one (1) governance workflow leveraging out of the box provided flows.
- Create, deploy and configure one (1) governance reporting database based on IBM Db2. Other database deployments will be considered on a case-by-case basis.
- Create and configure one (1) MDI and MDE project for data asset curation. This activity includes setting up the project security settings to assist with a compliance start for Client.
- Create and configure up to five (5) Data Protection Rules based on users, tags and groups and any installed/built business terms/classifications.
- Execute MDI and MDE jobs for the assets identified.
- Create and configure two (2) Data Asset catalogs.
- Create and configure one (1) data use and analysis project.

Data Quality (SaaS) (from "Build Data Quality (SaaS).pdf")

Service Description

IBM Technology Expert Labs Build Data Quality (SaaS)

This Service Description describes the Acceleration Services to support Client's use of an IBM software program. The applicable order documents provide pricing and additional details about Client's order. Upon acceptance of Client's order, this Service Description and the Cloud Services Agreement found at http://www.ibm.com/terms/csa (or equivalent agreement for Cloud Services between the parties) applies to Client's use of the Acceleration Services.

1. Acceleration Services

Acceleration Services are expert services, as described in this Service Description (SD) and are delivered remotely. For the purpose of this SD, if the base agreement references Cloud Services, then all such references apply to Acceleration Services, or Services, as applicable, and that base agreement shall govern the terms of this SD.

1.1 **Services**

The Client may select from the following available services.

1.1.1 IBM Technology Expert Labs Build Data Quality (SaaS)

This Service provides Client with a fundamental configuration of the Data Quality component of IBM watsonx.data intelligence. The fixed scope Service enables Client to initiate engagement. It provides capabilities to measure, monitor, and maintain data quality, helping to confirm that data used in AI and analytics use cases remain well curated and trustworthy by leveraging the IBM Knowledge Catalog.

The scope includes the following:

- Deploy a single cross-industry Knowledge Accelerator (KA)
- Define, create and customize one (1) governance workflow.
- Create five (5) Data Quality Definitions and up to 20 Data Quality rules.
- Create, configure and populate a Data Quality project.
- Configure Data Quality Service Level Agreement (SLA) workflows.

This offering includes a maximum of 120 hours of consultation and delivery with 12 hours of program management per Engagement and consists of the following Activities:

- Kick-off to set expectations for the engagement.
- Create and configure Cloud Pak for Data users, groups and roles specific to the Data Quality offering.
- Create and configure platform connections for two (2) existing Client data sources, with up to ten (10) tables total.
- Deploy with guidance on using one (1) KA, aligned with Client specific industry and/or requirements.

- Define, create and customize one (1) governance workflow configuration leveraging out of the box provided flows.
- Create, deploy and configure one (1) governance reporting database set based on IBM Db2. Other database deployments will be considered on a case-by-case basis.
- Create and configure one (1) MDI and MDE project for data asset curation. This activity includes setting up the project security settings to assist with a compliance start for Client.
- Execute MDI and MDE jobs for the data source assets connected in previous step.
- Create and configure one (1) project for Data Quality.
- Data Quality configuration automation approach.

4. IBM Technology Expert Labs Manta Data Lineage QuickStart SaaS (from "Manta Data Lineage QuickStart SaaS.pdf")

Service Description

IBM Technology Expert Labs Manta Data Lineage QuickStart SaaS

This Service Description describes the Acceleration Services to support Client's use of an IBM software program. The applicable order documents provide pricing and additional details about Client's order. Upon acceptance of Client's order, this Service Description and the Cloud Services Agreement found at http://www.ibm.com/terms/csa (or equivalent agreement for Cloud Services between the parties) applies to Client's use of the Acceleration Services.

1. Acceleration Services

Acceleration Services are expert services, as described in this Service Description (SD) and are delivered remotely. For the purpose of this SD, if the base agreement references Cloud Services, then all such references apply to Acceleration Services, or Services, as applicable, and that base agreement shall govern the terms of this SD.

1.1 **Services**

The Client may select from the following available services.

1.1.1 IBM Technology Expert Labs Manta Data Lineage QuickStart SaaS

The Service provides a fixed scope services to enable customers to get started quickly with Manta Data Lineage for a clear and comprehensive picture of their data landscape.

The scope includes the following:

- Up to five (5) supported scanner connections overall
- Maximum of one Extract, Transform, Load (ETL) scanner
- Ingestion maximum of up to 20K scripts
- Lineage Review of up to 3 data sources

This offering includes a maximum of 104 hours of consultation and delivery with 21 hours of program management per Engagement and consists of the following Activities:

- Kickoff to set expectations for the engagement
- Remote Assessment to review prerequisites necessary to deploy the Solution
- Discovery Workshop to understand requirements, identify use case and agree on scanners to implement
- Installation of Manta software or service on a supported platform
- Configure and validate agreed upon scanners / connections
- Metadata import for up to 2 supported data source systems into IKC (Cloud Pak for Data only)
- Lineage Review with Client SME to demonstrate how to leverage the lineage to meet use case objectives
- Operational workshop to provide Client with understanding of the process for scanner configuration, validation and basic operation

2. Data Processing and Protection Data Sheets

IBM's Data Processing Addendum at http://ibm.com/dpa (DPA) and the Data Processing and Protection Data Sheet(s) (referred to as data sheet(s) or DPA Exhibit(s)) in the links below applies to IBM's processing of personal data on behalf of Client.

http://www.ibm.com/terms/?id=DPA-EX_SW_Sces

3. Service Levels and Technical Support

No Service Level Agreement or Technical Support is available for this Service Description.

4. Charges

4.1 **Charge Metrics**

The Acceleration Service is available under the charge metric specified in the Transaction Document:

5. IBM Expert Labs Data Platform Accelerator Services – SaaS (from "Data Platform Accelerator Services – SaaS.pdf")

Service Description

IBM Expert Labs Data Platform Accelerator Services – SaaS

Upon acceptance of Client's order, this Service Description applies to the Acceleration Services to support Client's Cloud Service. The applicable order documents provide pricing and additional details about Client's order.

1. Acceleration Service

Acceleration Services are expert services, as described in this SD, and are provided remotely to the Client. For the purpose of this SD, if the base agreement references Cloud Services, then all such references apply to Acceleration Services, as applicable, and that base agreement shall govern the terms of this SD. These Acceleration Services are available for each of the following technology areas:

- watsonx.ai

- watsonx.data
- watsonx.governance & OpenPages
- Watson Orchestrate & AI Assistants
- Data Fabric
- watsonx.data integration
- watsonx.data intelligence
- Business Analytics & Cognos
- Content Management
- Workflow Automation
- Databases
- Cloud Pak for Data
- Cloud Pak for Business Automation

1.1 **Services**

The Client may select from the following available services.

1.1.1 IBM Expert Labs One Day Consulting Services

This Service provides a maximum of 8 hours of consultation and delivery per Engagement and consists of the following Activities:

Activity 1 – Conduct Project Kickoff Meeting

IBM will conduct a kick-off meeting on a mutually agreed to date at the commencement of this consulting service to perform the following activities such as:

- Review Client requirements;
- Understand Client objectives;
- Define Client and IBM roles and responsibilities; and
- Document planned activities, priorities and timelines.

Activity 2 – Consulting Services

As prioritized and defined in Activity 1, and as time permits, IBM will spend the remainder of the hours to perform activities such as:

- Use case identification and validation;
- Delivery proposal development;
- Defining or reviewing architecture;
- Defining business value;
- Defining migration approach;
- Client data assessment;

6. Service Description Build Data Quality from “IBM Technology Expert Labs Build Data Quality”

- This Service provides Client with a fundamental configuration of the Data Quality component of IBM watsonx.data intelligence. The fixed scope Service enables Client to initiate engagement. It provides capabilities to measure, monitor, and maintain data quality, helping to confirm that

data used in AI and analytics use cases remain well curated and trustworthy by leveraging the IBM Knowledge Catalog. The scope includes the following:

- Deploy a single cross-industry Knowledge Accelerator (KA)
- Define, create and customize one (1) governance workflow.
- Create five (5) Data Quality Definitions and up to 20 Data Quality rules.
- reate, configure and populate a Data Quality project.
- Configure Data Quality Service Level Agreement (SLA) workflows. This offering includes a maximum of 120 hours of consultation and delivery with 12 hours of program management per Engagement and consists of the following Activities:
- Kick-off to set expectations for the engagement.
- Create and configure Cloud Pak for Data users, groups and roles specific to the Data Quality offering.
- Create and configure platform connections for two (2) existing Client data sources, with up to ten (10) tables total.
- Deploy with guidance on using one (1) KA, aligned with Client specific industry and/or requirements.
- Define, create and customize one (1) governance workflow configuration leveraging out of the box provided flows.
- Create, deploy and configure one (1) governance reporting database set based on IBM Db2. Other database deployments will be considered on a case-by-case basis.
- Create and configure one (1) MDI and MDE project for data asset curation. This activity includes setting up the project security settings to assist with a compliance start for Client.
- Execute MDI and MDE jobs for the data source assets connected in previous step.
- Create and configure one (1) project for Data Quality.
- Data Quality configuration automation approach.

7. Scope Document for customer and Build AI Use Case Risk Management with watsonx.governance

Scope Document for customer and Build AI Use Case Risk Management with watsonx.governance

Scope of Services

Estimated Schedule

IBM will perform the following activities over the course of a 8 week period with a scheduled start date mutually agreed upon by both IBM and Client.

Work Activities

1. **Solution Planning & Design** a. Enable customer on AI use case onboarding framework using watsonx.governance b. Identify and review prioritized AI use cases in scope c. Jointly map customer's business process for onboarding new or existing AI use cases and align to a RACI. d. Define import process for populating existing AI use cases into watsonx.governance use case inventory e.g. IBM Fastmap tooling (integration with 3rd party tools or systems is not included in this scope of work). e. Document and review the AI use case onboarding solution design with the client team. f. If required, provide up to one revision of the solution design document to incorporate Client feedback.
2. **Solution Configuration** For a single watsonx.governance environment: a. Configure out of the box AI use case approval workflow to aligned business process with up to 3 additional review/approval stages. b. Create a business entity hierarchy (up to 5 business entities) in IBM watsonx.governance console. c. Configure a sample risk and/or applicability assessment questionnaire leveraging Governance Console out of the box assessments (e.g. EU AI Act applicability assessment) & out of the box risk library. d. Populate the inventory with up to 2 AI use cases (new or existing) based on agreed onboarding process (no integration with 3rd party tools or systems). e. Configure Governance Console OOTB user roles and permissions for governing the 2 AI use cases. (Configuration of custom user roles is out of scope). f. Configure Governance console to include up to 5 custom fields. g. Configure up to 4 views in Governance Console to include custom fields. h. Demonstrate automated risk rating evaluations using configured risk assessment. i. Configure dashboard in Governance Console with up to 5 additional widgets to provide business view of AI use cases (e.g. use case by approval status, by deployment status, by BU, by risk level etc.). j. Provide up to 16 hours of knowledge sharing to customer covering: k. IBM watsonx.governance out of the box Dashboard and how to configure. l. AI use case creation in watsonx.governance, risk classification and approval workflow m. Assessment questionnaire configuration n. Risk rating configuration based on questionnaire o. Add custom Risks to the library
3. **Activation** a. Promote the configuration from the 1st watsonx.governance environment to a 2nd watsonx.governance environment. b. Provide post-promotion assistance for the 2nd environment for up to 5 days immediately following application of the configuration.
4. **Project Close Out** a. Review engagement activities. b. Provide a recommendation for next steps.

Client Responsibilities

- Work with IBM to identify Client owners for key activities, business and technical knowledge and other areas as requested.
- Make appropriate personnel available to assist IBM in the performance of its responsibilities
- Provide IBM with access to the technical and business documentation and data reasonably necessary to perform the activities defined in this engagement.
- Provide IBM with feedback, decisions and information reasonably necessary to perform the activities defined in this engagement.
- On-board IBM personnel, if required
- Provide a deployed instance of OpenPages
- Provide IBM with the logistics to access the environment.
- Ensure all Client security approvals are obtained.

2025 IBM CORPORATION | CLIENT AND IBM CONFIDENTIAL

Assumptions

- All IBM services are delivered remotely;
- Normal work hours will be 8:30am – 5:00pm, Monday through Friday, in the local time of the location the work will be performed;
- The scope is limited to configuration of a single environment configuration, single use case.
- Client has appropriate software already deployed in a SaaS or non-SaaS environment of sufficient size for services engagement.

9. Scope Documents - Enterprise Search (from "Scope Documents - Enterprise Search - all sizes - 20250918_163133.docx")

Table of Contents (Links)

Scope Document - Enterprise Search - Small (SaaS)

Work Activities

IBM will perform the following activities over the course of the entire project. The Services are comprised of the following activities to be delivered over a 7 week period.

- DESIGN: Conduct a kick-off meeting to align expectations for the engagement, to validate all pre-requisites are in place, and align on expected timeline and effort.
- DESIGN: Conduct architecture and design meetings to plan the generative search. This activity is limited to 40 hours.
- DEVELOP: Perform search result tuning. This activity is limited to 120 hours.
- QUALIFY: Assist the client with performing User Acceptance Testing. This activity is limited to 40 hours.
- ROLLOUT: Advise the client on Watson Assistant chat UI or API integration into the client's systems. An example chat / query UI will be provided.
- ROLLOUT: Conduct knowledge transfer specific to the application, environment, and architecture to support the application going live. This activity is limited to 40 hours.

Client Responsibilities

- DESIGN: Provide access to content to be indexed watsonx Discovery for retrieval.
- DESIGN: Work with IBM to identify client owners for key activities, business and technical knowledge and other areas, as requested.
- DEVELOP: Provide necessary resources as needed to support or troubleshoot
- DEVELOP: The client will provide API access to an existing configured search engine.
- QUALIFY: Perform User Acceptance Testing.
- QUALIFY: Verify that approvals are in place for the application to go-live.
- ROLLOUT: Integration of a chat UI into client systems will be performed by the client.
- ROLLOUT: Participate in the technical knowledge transfer session with IBM.

Assumptions

- All apps or data sources used will be reachable via the public internet.
- All credentials and connectivity required will be available at the start of the project.
- If IBM identifies areas of remediation during the architecture, design, code, and environment workshops, an additional SOW will be required to perform the required changes.
- If the application fails User Acceptance testing and remediation is required, an additional SOW will be required to perform the required changes.
- If travel and living is deemed necessary, to complete this DoU. those charges are in addition to the quoted services charges and... (truncated in original).

10. Apptio Jumpstart Service: Application Costing with Infrastructure Detail (from "service-description-apptio-jumpstart-application-costing-infrastructure.pdf")

Apptio Jumpstart Service: Overview

Apptio will perform the Professional Services described in the Application Costing with Infrastructure Detail package set forth herein to configure the Cost Transparency solution. The Professional Services will follow an iterative, workshop-based approach to implement the configured functionality set forth in Table 1:

Table 1

Configuration Iteration Name	Configured Functionality	Associated Apptio Product Module
Application Labor and Projects	Review and manage labor spend and resources tied to Application Development and Support	Cost Transparency (Foundation, and Apps & Services)
	Understand Labor and Project spend by App	Cost Transparency (Foundation, and Apps & Services)
Application Assets and Vendors	Review and manage Vendor and Contract spend, including impact per Application	Cost Transparency (Foundation, and Apps & Services)
	Understand 'Direct' applications costs, leveraging General Ledger	Cost Transparency (Foundation, and Apps & Services)
	Manage and rationalize Applications leveraging visibility of Addressable/Non-Addressable spend	Cost Transparency (Foundation, and Apps & Services)
Application Infrastructure Analysis	Visibility into Labor, Asset, Project and Vendor spend for Servers, Storage, and Data Center	Cost Transparency Apps & Services
	Ability to rationalize apps with Total Cost of Ownership	Cost Transparency Apps & Services

Project Tasks

Apptio's approach includes three (3) phases: Discover & Plan, Configure & Rollout, and Project Closeout.

This consolidated document captures provided service descriptions. Documents have been abbreviated for clarity.

IBM W9

Form **W-9**
(Rev. March 2024)
Department of the Treasury
Internal Revenue Service

Request for Taxpayer Identification Number and Certification

Go to *www.irs.gov/FormW9* for instructions and the latest information.

Give form to the requester. Do not send to the IRS.

Before you begin. For guidance related to the purpose of Form W-9, see *Purpose of Form*, below.

Print or type. See *Specific Instructions* on page 3.

1 Name of entity/individual. An entry is required. (For a sole proprietor or disregarded entity, enter the owner's name on line 1, and enter the business/disregarded entity's name on line 2.)

International Business Machines Corporation

2 Business name/disregarded entity name, if different from above.

IBM Corporation

3a Check the appropriate box for federal tax classification of the entity/individual whose name is entered on line 1. Check only **one** of the following seven boxes.

☐ Individual/sole proprietor ☑ C corporation ☐ S corporation ☐ Partnership ☐ Trust/estate

☐ LLC. Enter the tax classification (C = C corporation, S = S corporation, P = Partnership) . . . .

Note: Check the "LLC" box above and, in the entry space, enter the appropriate code (C, S, or P) for the tax classification of the LLC, unless it is a disregarded entity. A disregarded entity should instead check the appropriate box for the tax classification of its owner.

☐ Other (see instructions)

4 Exemptions (codes apply only to certain entities, not individuals; see instructions on page 3):

Exempt payee code (if any) 5

Exemption from Foreign Account Tax Compliance Act (FATCA) reporting code (if any)

(Applies to accounts maintained outside the United States.)

3b If on line 3a you checked "Partnership" or "Trust/estate," or checked "LLC" and entered "P" as its tax classification, and you are providing this form to a partnership, trust, or estate in which you have an ownership interest, check this box if you have any foreign partners, owners, or beneficiaries. See instructions . . . . . ☐

5 Address (number, street, and apt. or suite no.). See instructions.

1 North Castle Drive, c/o Income Tax Department

Requester's name and address (optional)

6 City, state, and ZIP code

Armonk, NY, 10504

7 List account number(s) here (optional)

Part I Taxpayer Identification Number (TIN)

Enter your TIN in the appropriate box. The TIN provided must match the name given on line 1 to avoid backup withholding. For individuals, this is generally your social security number (SSN). However, for a resident alien, sole proprietor, or disregarded entity, see the instructions for Part I, later. For other entities, it is your employer identification number (EIN). If you do not have a number, see *How to get a TIN*, later.

Note: If the account is in more than one name, see the instructions for line 1. See also *What Name and Number To Give the Requester* for guidelines on whose number to enter.

Social security number

			–			–				

or

Employer identification number

1	3	–	0	8	7	1	9	8	5

Part II Certification

Under penalties of perjury, I certify that:

1. The number shown on this form is my correct taxpayer identification number (or I am waiting for a number to be issued to me); and
2. I am not subject to backup withholding because (a) I am exempt from backup withholding, or (b) I have not been notified by the Internal Revenue Service (IRS) that I am subject to backup withholding as a result of a failure to report all interest or dividends, or (c) the IRS has notified me that I am no longer subject to backup withholding; and
3. I am a U.S. citizen or other U.S. person (defined below); and
4. The FATCA code(s) entered on this form (if any) indicating that I am exempt from FATCA reporting is correct.

Certification instructions. You must cross out item 2 above if you have been notified by the IRS that you are currently subject to backup withholding because you have failed to report all interest and dividends on your tax return. For real estate transactions, item 2 does not apply. For mortgage interest paid, acquisition or abandonment of secured property, cancellation of debt, contributions to an individual retirement arrangement (IRA), and, generally, payments other than interest and dividends, you are not required to sign the certification, but you must provide your correct TIN. See the instructions for Part II, later.

Sign Here Signature of U.S. person Theresa Glendon Date 1/3/2025

General Instructions

Section references are to the Internal Revenue Code unless otherwise noted.

Future developments. For the latest information about developments related to Form W-9 and its instructions, such as legislation enacted after they were published, go to *www.irs.gov/FormW9*.

What's New

Line 3a has been modified to clarify how a disregarded entity completes this line. An LLC that is a disregarded entity should check the appropriate box for the tax classification of its owner. Otherwise, it should check the "LLC" box and enter its appropriate tax classification.

New line 3b has been added to this form. A flow-through entity is required to complete this line to indicate that it has direct or indirect foreign partners, owners, or beneficiaries when it provides the Form W-9 to another flow-through entity in which it has an ownership interest. This change is intended to provide a flow-through entity with information regarding the status of its indirect foreign partners, owners, or beneficiaries, so that it can satisfy any applicable reporting requirements. For example, a partnership that has any indirect foreign partners may be required to complete Schedules K-2 and K-3. See the Partnership Instructions for Schedules K-2 and K-3 (Form 1065).

Purpose of Form

An individual or entity (Form W-9 requester) who is required to file an information return with the IRS is giving you this form because they

Cat. No. 10231X Form **W-9** (Rev. 3-2024)